MISCELLANEOUS

A case study in applying artificial intelligence-based named entity recognition to develop an automated ophthalmic disease registry

Carmelo Z Macri[1,2] · Sheng Chieh Teoh[2] · Stephen Bacchi[1,2] · Ian Tan[2] · Robert Casson[1,2] · Michelle T Sun[1,2] · Dinesh Selva[1,2] · WengOnn Chan[1,2]

Abstract

Purpose Advances in artificial intelligence (AI)-based named entity extraction (NER) have improved the ability to extract diagnostic entities from unstructured, narrative, free-text data in electronic health records. However, there is a lack of ready-to-use tools and workflows to encourage the use among clinicians who often lack experience and training in AI. We sought to demonstrate a case study for developing an automated registry of ophthalmic diseases accompanied by a ready-to-use low-code tool for clinicians.

Methods We extracted deidentified electronic clinical records from a single centre's adult outpatient ophthalmology clinic from November 2019 to May 2022. We used a low-code annotation software tool (Prodigy) to annotate diagnoses and train a bespoke spaCy NER model to extract diagnoses and create an ophthalmic disease registry.

Results A total of 123,194 diagnostic entities were extracted from 33,455 clinical records. After decapitalisation and removal of non-alphanumeric characters, there were 5070 distinct extracted diagnostic entities. The NER model achieved a precision of 0.8157, recall of 0.8099, and F score of 0.8128.

Conclusion We presented a case study using low-code artificial intelligence-based NLP tools to produce an automated ophthalmic disease registry. The workflow created a NER model with a moderate overall ability to extract diagnoses from free-text electronic clinical records. We have produced a ready-to-use tool for clinicians to implement this low-code workflow in their institutions and encourage the uptake of artificial intelligence methods for case finding in electronic health records.

Key messages

What is known:

● Artificial intelligence-based natural language processing has significantly advanced the ability to extract information from clinical free-text.

● Despite recent advances, implementation of these techniques for clinicians is limited by experience, training, and access to ready-to-use tools.

What is new:

● Through a case study, we demonstrate the potential of artificial intelligence-based named entity recognition to extract diagnostic entities from clinical free-text electronic health records to create an automated registry of ophthalmic diseases.

● This case study is accompanied by a low-code, ready-to-use tool to reproduce the registry in any institution.

✉ Carmelo Z Macri
    carmelo.macri@adelaide.edu.au

1   Discipline of Ophthalmology and Visual Sciences, The
    University of Adelaide, Adelaide, South Australia, Australia

2   Department of Ophthalmology, The Royal Adelaide Hospital,
    Adelaide, South Australia, Australia

Introduction

Artificial intelligence-based natural language processing (NLP) techniques have significantly improved the ability to extract information from free text [1]. This technology has important implications for improving the recording of diagnoses in electronic health records. Supplementing manually coded diagnoses with those found in text improves patient cohort identification in studies involving the secondary use of electronic health records [2]. However, applying new and advanced artificial intelligence methods for diagnostic named entity recognition (NER) requires expert knowledge of these techniques and the skills to implement them. These skills are unfamiliar to most clinicians and are a significant barrier to implementing NLP in clinical and research workflows.

Artificial intelligence-based methods have advantages over the previous dictionary and rule-based techniques for clinical named entity recognition. Dictionary-based approaches such as the clinical Text Analysis and Knowledge Extraction System (cTAKES) are early examples of NLP that provided good performance for NER with clinical text. The cTAKES algorithm implemented terminology-agnostic dictionary look-up within a noun-phrase look-up window [3]. However, dictionary-based approaches are limited by the uniqueness of biomedical vocabulary, including abbreviations [4–6], misspellings [7], variable representations of similar concepts [8], ambiguity [9], and variable representations of numbers in text [10]. Rule-based approaches can often achieve high performance [11, 12], but are practically limited by needing to be tailored to specific entities and texts, which restricts generalizability, and are resource intensive requiring extensive expert knowledge and time to develop. Powerful feature-engineered supervised machine learning methods such as conditional random fields (CRF) and support vector machine algorithms further improved the performance of NER beyond dictionary and rule-based approaches, demonstrating the potential application of machine learning to natural language processing and increasing their use [13]. Deep learning methods, including neural networks, have shown additional increases in performance [14, 15]. In particular, recurrent neural networks have shown examples of superior performance to CRF for clinical text [1]. More recent advancements in transfer learning and transformer-based models have improved performance even further [15]. Artificial intelligence offers more generalisable approaches to disease identification without extensive clinician input.

Despite a general awareness of the uses of artificial intelligence, clinicians' lack of artificial intelligence training and experience may present a barrier to implementing such technology [16–18]. Education of clinicians regarding artificial intelligence and assistance with implementation is an emerging priority [19], given that clinicians will be a critical factor in adopting AI in healthcare. Developing artificial intelligence-based tools and workflows that are easy to use, production-ready, and low-code may assist in facilitating the introduction of artificial intelligence techniques into healthcare and research. There are few ready-to-use tools to apply to clinical text for diagnostic registry production using clinical NER [15]. Thus, we sought to develop and demonstrate the application of low-code artificial intelligence-based NLP tools applied to electronic clinical records to build an automated registry of ophthalmic diseases.

Methods

We performed this study at the Royal Adelaide Hospital, Adelaide, Australia, with the approval of the institutional Human Research Ethics Committee, adhering to the tenants of the Declaration of Helsinki. We extracted deidentified free-text ophthalmology clinic records from the EHR system for all adult outpatient ophthalmology clinics between November 2019 and May 2022. All notes were free text and written in English.

We performed dataset annotation and NER model training using a low-code annotation software tool (*Prodigy*, ExplosionAI GmbH, Berlin, Germany) [20]. Prodigy is an active learning-based annotation tool and integrates with the spaCy natural language processing learning library. The architecture of the spaCy model is not open source but is described as using sub-word features, Bloom embeddings, and a deep convolutional neural network with residual connections. The tool enables the annotation of diagnoses by highlighting text in a graphical user interface displayed in a web browser (Fig. 1) [21]. The tool uses simple, one-line text commands entered into the terminal to execute tasks. These tasks are pre-scripted Python functions that initialise dataset annotation and train NER models. Figure 2 summarises the workflow.

Annotation was performed by a single qualified medical practitioner investigator with graduate ophthalmic experience (CM). Only ophthalmic diagnostic entities were annotated (Fig. 1). Non-ophthalmic diagnoses listed in

Fig. 1 The graphical user interface for annotation dataset creation with example annotations of diagnostic entities

past medical history when this occurred were not annotated. The spans of words containing the complete description of the diagnosis were annotated to ensure extractions were interpretable, non-ambiguous, and preserved a contextual window on either side of the diagnosis.

The annotation command tokenises the electronic clinical records into words to prevent errors of partial selection when annotating. Using the graphical user interface, we annotated the first 1000 health records to create an initial dataset of annotations (Fig. 2). We annotated only words relevant to the diagnosis, annotating multiple-word diagnoses as a complete annotation. Using the initial annotation dataset, we trained an initial NER model, which we subsequently used to provide suggested annotations in further dataset annotation to increase annotation efficiency.

A further and larger annotation dataset was created by annotating a proportion of the remaining clinical records and correcting the suggestions made by the initial NER model. We included only new records not previously annotated to create this dataset. We calculated accuracy statistics at approximately 500 note intervals by training a model using increasing proportions (25%, 50%, 75%, 100%) of the total annotations. Annotation of the clinical records continued until model accuracy showed minimal-to-no further improvement within the last 25%, occurring at 1923 records.

Using the low-code tool, we trained a final NER model using both the initial and larger annotation datasets. The model evaluation metrics included precision, recall, and standard F score [22]. The model training command reserves a proportion of annotations to evaluate the model and produce accuracy statistics after training. Therefore, creating a separate gold standard evaluation dataset is not required to evaluate the model's

Fig. 2 Summary of the work-flow to build the low-code automated ophthalmic disease registry
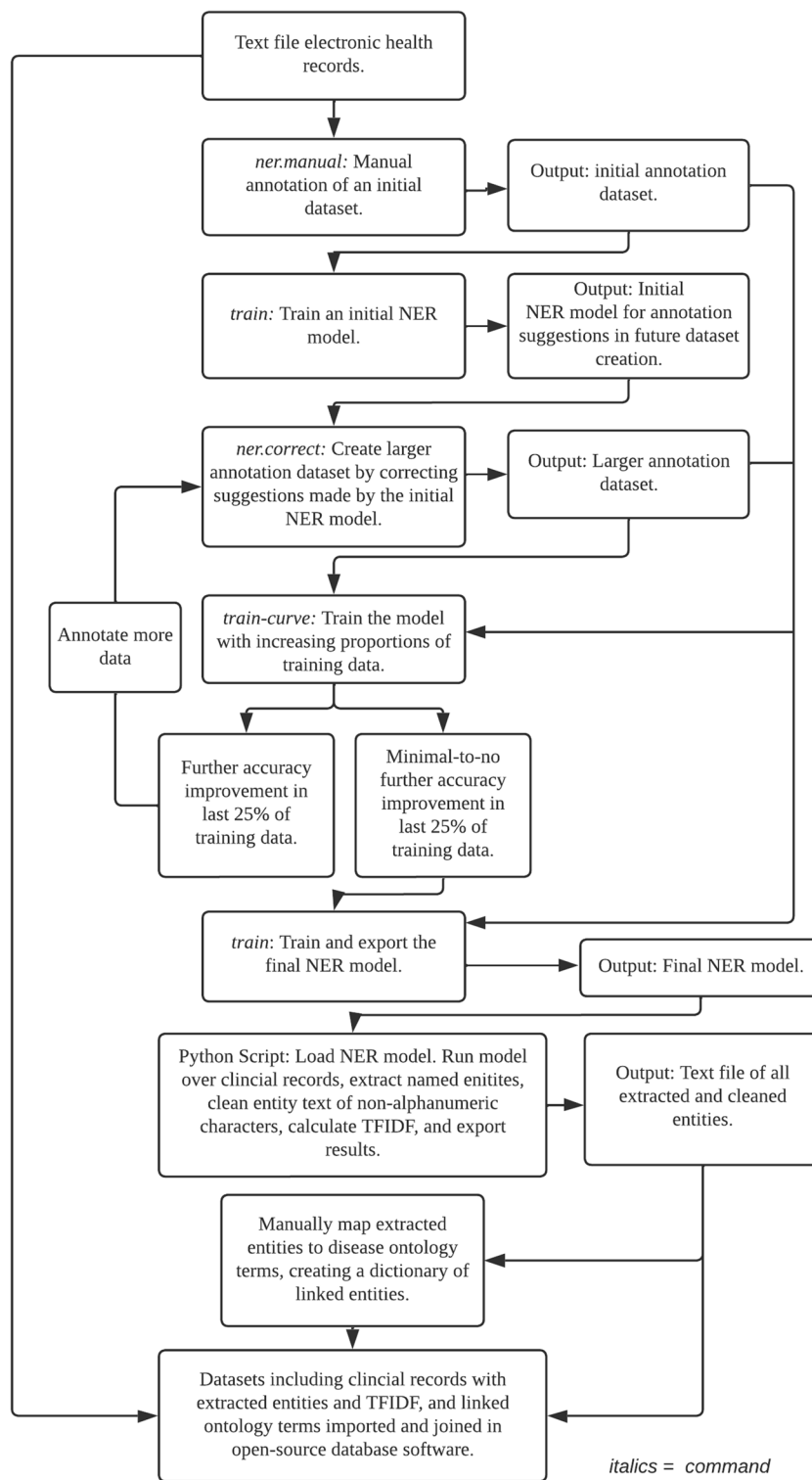
performance. We used 20% of the annotations to produce the precision, recall, and *F* score. Precision refers to the ratio of true positives to the sum of true and false positives (TP/TP + FP), and recall refers to the ratio of true positives to the sum of true positives and false negatives (TP/TP + FN). NER model errors were analysed by the

proportion of complete false positives, complete false negatives, and right label with overlapping span, as presented by Nejadgholi et al. [23].

To extract the diagnostic entities, we used the *spaCy* (v3.1.4) library to load and run the model over the entire set of clinical records. After extraction, regular

expressions cleaned the entities to remove capitalisation and non-alphanumeric characters. In addition, we used the *gensim* (v4.1.2) library to calculate the term frequency-inverse document frequency (TF-IDF) for each entity-document pair to include for use in the registry. A binary weight was used for the term frequency and pivoted unique normalisation for document length normalisation. We used a binary weight as only the appearance of the entity in the document was relevant. Pivoted unique normalisation was used to counter bias introduced by document length and align the probabilities of retrieval and relevance [24], given that clinical notes can vary in length.

We manually mapped a proportion of extracted entities representing common terms to SNOMED-CT (International Edition, version 2021-07-31) terms and corresponding codes. The datasets, including the clinical records, extracted entities, and their mapped SNOMED-CT terms, were imported into a free and open-source database management tool (*Metabase*, San Francisco, CA, USA) [25]. Datasets were joined via common data elements to produce a final registry containing patient

medical record numbers, health records, extracted entities, and linked SNOMED-CT terms (Fig. 3).

We have condensed the steps for creating this registry into a series of sequential batch files (text files that execute a sequence of commands) for simple reproduction in any institution. Users must supply their electronic records to build the registry using our pre-trained NER model. Alternatively, users can train an institution-specific NER model in place of this using a variety of the available low-code annotation tools [26]. The reproducible registry files are hosted on GitHub (https://github.com/OphRL/AutoRegistry) along with instructions.

Results

The model achieved an *F* score of 0.8128, precision (ratio of true positives to the sum of true positives and false positives) of 0.8157, and recall (ratio of true positives to the sum of true positives and false negatives) of 0.8099. The model was run over 33,455 notes, and a

Diagnosis Registry Database

StudyID	Extracted Entity	SNOMED Description	SNOMED ID	TFIDF	Clinical Record
11131191	rd	Retinal detachment (disorder)	42059000	0.7	# L endogenous en
11131191	pco	Secondary cataract (disorder)	766834007	1.1	# L endogenous en
11129859	crao	Central retinal artery occlusion (disorder)	38742007	2.45	Clinical Details　1
11129859	hypertensive uveitis	Anterior uveitis (disorder)	410692006	2.86	Clinical Details　1
11129526	uveitis	Uveitis (disorder)	128473001	2.15	Emergency Ophth
11129526	pvd	Posterior vitreous detachment (disorder)	247081001	1.67	Emergency Ophth
11129526	uveitis	Uveitis (disorder)	128473001	2.15	Emergency Ophth
11129193	keratitis	Keratitis (disorder)	5888003	2.32	Clinical Details　#
11129193	microbial keratitis	Keratitis caused by infection (disorder)	733312003	2.07	Clinical Details　#
11128527	hypotony	Hypotony of eye (disorder)	19721008	3.23	dmf, hypotony. rev
11127861	posterior uveitis	Posterior uveitis (disorder)	43363007	3.26	LE posterior uveiti
11127861	vitreous condensation	Vitreous liquefaction (disorder)	50801008	3.19	LE posterior uveiti
11127861	erm	Epiretinal membrane (disorder)	367649002	1.26	LE posterior uveiti
11127195	pdr	Proliferative retinopathy due to diabetes mellitus (disorder)	59276001	1.43	Clinical Details　E
11127195	pseudophakic	Pseudophakia (disorder)	95217000	1.18	Clinical Details　E
11126862	vh	Vitreous hemorrhage (disorder)	31341008	1.71	improving. mild VI

Fig. 3 Interface of the automated ophthalmic disease registry

Table 1 Most frequent entities extracted from text (decapitalised and non-alphanumeric characters removed)

Extracted entity	Number	Proportion of total entities (%)
cataract	6419	5.2
ppv	3744	3.0
erm	3476	2.8
rd	2887	2.3
pseudophakic	2727	2.2
cataracts	2533	2.1
iol	2296	1.9
phaco	2240	1.8
cmo	1956	1.6
poag	1940	1.6
pdr	1918	1.6
vh	1893	1.5
glaucoma	1746	1.4
pvd	1592	1.3
trab	1385	1.1
avastin	1382	1.1
pterygium	1367	1.1
dmo	1284	1.0
cnvm	1256	1.0
prp	1204	1.0

total of 123,194 named entities were extracted, 5070 of which were distinct (after decapitalisation and removing non-alphanumeric characters). The most frequently extracted diagnostic entities included 'cataract' (5.2%), followed by 'ppv' (3.0%), 'erm' (2.8%), 'rd' (2.3%), and 'pseudophakic' (2.2%). The 20 most frequent extractions are presented in Table 1.

There were 159 type one (complete false positives), 102 type two (complete false negatives), and 20 type five (right label, overlapping span) mismatches. Figure 4 illustrates an example of a note containing correctly predicted diagnostic entities (yellow), false negatives (red), and false positives (green). The figure shows the correct labelling of 'optic neuropathy'. However, the model did not predict the diagnostic entity 'atypical optic neuritis', resulting in a false negative. In addition, the model predicted the listed differential 'GCA' as a diagnostic entity which was recorded as a false positive.

Table 2 shows examples of lexical representations of cranial nerve palsies in the clinical records. The entities exemplify misspellings, abbreviations, acronyms, varying forms for the same concept, variable representation of numbers using words, and Arabic and Roman numerals.

Fig. 4 Example of clinical record with the NER model applied showing correct predictions, false negatives, and false positives. Yellow: correctly predicted diagnostic entities. Red: false negatives. Green: false positives

Discussion

Using a low-code workflow, we trained a NER model with moderate precision (0.8157) and recall (0.8099) and overall performance (F score 0.8128) in extracting diagnoses from free-text clinical records. Most errors were due to false positives, followed closely by false negatives. Overlapping spans accounted for a small proportion (7.1%) of errors during evaluation. A higher false positive rate is unlikely to impact the functioning of an automated registry, given that

Table 2 Examples of the various lexical representations of cranial nerve palsies in ophthalmic clinical records (decapitalised and non-alphanumeric characters removed)

Concept	Entities
Cranial nerve palsy	cn palsy, craneal nerve palsy, cranial nerve palsy
3rd cranial nerve palsy	3rd cn palsy, 3rd nerve palsy, cn iii microvascular palsy, cn iii palsy, cn3 palsy, cn3fourth palsy, cniii palsy, iii cn palsy, iii n palsy, iii nerve palsy, microvascular third nerve palsy, third nerve palsy, third nerve palsy suspect, total cn3 palsy
4th cranial nerve palsy	cn 4 palsy, cn 4th palsy, cn iv palsy, cn3fourth palsy, cn4 palsy, cniv palsy, congenital cn4 palsy, forth nerve palsy, fourht nerve palsy, fourth n palsy, fourth nerve palsy, fourth nerve paresis, iv cn palsy, iv n palsy, iv nerve palsy, iv palsy
5th cranial nerve palsy	cn v palsy, cn5 palsy, trigeminal nerve palsy
6th cranial nerve palsy	6th nerve palsy, 6th palsy, abducens nerve palsy, abducens palsy, acute cn vi palsy, cn 6 palsy, cn 6th palsy, cn vi palsy, cn6 new palsy, cn6 palsy, cnvi palsy, cranial nerve vi palsy, traumatic cn vi palsy, vi and vii palsy, vi cn, vi cn palsy, vi cranial nerve palsy, vi n palsy, vi n paresis, vi nerve palsy, vi nerve paresis, vi palsy, vith cnp, vith cranial nerve palsy, vith nerve palsy
7th cranial nerve palsy	bell's palsy, bells palsy, branch viin palsy, cn 7 palsy, cn vii, cn vii palsy, cnvii palsy, facial n palsy, facial nerve deficit, facial nerve palsies, facial nerve palsy, facial nerve paralysis, facial nerve static palsy, facial nerve weakness, facial palsy, facial vii palsy, parotid gland resection cn 7th palsy, total facial nerve palsy, vi and vii palsy, vii palsy, viith palsy

the aim is to detect all possible diagnoses present. However, false negatives are an area of potential improvement. The false positive pictured in Fig. 4 shows an example of a prediction that was incorrect due to its context rather than an incorrect diagnostic entity. Given that differential lists are a common occurrence, this may contribute to the higher false positive rate.

The complexities of clinical natural language are demonstrated through examples of variable representations of cranial nerve palsies in free text (Table 2). These entity examples illustrate the presence of misspellings, abbreviations, acronyms, variable forms of similar concepts, and variable representations of numerical expressions in ophthalmic notes. Low-code NLP tools enable the rapid creation of a disease registry containing a broad range of diagnoses in free-text electronic clinical records without requiring extensive clinician input. We implemented this pipeline in a ready-to-use tool to implement this workflow in any institution to create a disease registry.

Low-code NLP tools aim to reduce the barriers to implementing new and advanced artificial intelligence-based techniques for entity recognition in clinical and research workflows. We performed annotation using a user-friendly graphical interface, which was initialised using simple commands in the terminal (the text-based interface which enables interaction with the computer's files and directories). Given that annotated datasets are required for supervised learning techniques, an increasing number of annotation tools are now available to create these datasets efficiently [26]. Features such as annotation suggestions are important, given that pre-annotation has previously been shown to improve annotation speed [27].

Rule-based approaches to extracting entities may perform well in task and domain-specific applications but are time-consuming and task-specific and require significant domain expert input when compared. Previous applications of such techniques to disease registries have included the use of regular expressions (text pattern matching) [28], modified tools based on regular expressions [29], and NLP tools using pre-trained models augmented with rule-based techniques [30, 31] [32]. Matching entities through regular expressions requires intimate knowledge of the representation of entities in clinical text and pre-specification of the patterns to detect. This specification is time-consuming and inflexible. For example, designing regular expressions to detect all possible representations of cranial nerve palsies, as depicted in Table 2, is complex. There have been significant advancements in artificial intelligence-based techniques for clinical NER, particularly with the introduction of transfer learning and transformer-based models [15]. For example, Moquarrab et al. presented a novel deep learning-based technique to extract clinical entities from clinical notes in the i2b2 NLP challenge datasets [33]. The authors used a combination of techniques, including a convolutional neural network, bidirectional long short-term memory (Bi-LSTM), and conditional random fields with non-complex embeddings. They achieved an $F1$ score of 93.57 and 86.11 across the 2010 and 2012 i2b2 datasets, respectively, showing significant improvements above previous applications. For comparison, the combination of the Bi-LSTM model and bidirectional encoder representations from transformers (BERT) embeddings achieved an $F1$ score of 90.25 and 80.91 in the i2b2 2010 and 2012 datasets, respectively [34]. Other popular models for NER, such as the conditional random field, achieved an $F1$ score of 84.30 in the i2bs 2010

dataset [35]. While it is difficult to perform comparisons across studies due to differences in pre-processing, dataset, and methodological differences, the benefits and improving performance of artificial intelligence-based techniques for clinical NER are promising for applications in automated registry production. However, few tools ready for implementation are currently available [15].

An ophthalmic disease registry could play an important role in identifying and monitoring rare diseases through electronic health records. It is estimated that 263–446 million persons are affected by rare diseases globally at any time [36]. Despite the clear burden of rare diseases and the need for research, rare disease research is limited by recruitment and sample size issues [37]. Searching diagnostic codes for instances of rare diseases is restricted by underrepresentation in most common ontologies such as the International Classification of Diseases [38] [39]. Electronic health records have been used previously to identify rare diseases [40, 41]; however, approaches to detection relied on regular expressions [42, 43]. A NER registry approach eliminates the pre-specification of expressions and is not diagnosis-specific, allowing flexibility in the range of diseases to be monitored. DeLozier et al. previously developed a system to monitor rare diseases through electronic health records [43]. An email alert system was used to prompt investigators to review rare drug reactions in clinical notes to improve recruitment in prospective clinical trials of drug-induced torsades de pointes and Stevens-Johnson Syndrome and toxic epidermal necrolysis. The alert system increased the rate of recruitment and reduced the time to enrolment in the studies. Monitoring diseases in free-text fields via integration with alerting systems can improve the monitoring of rare diseases and reduce barriers to cohort identification for research.

Diagnoses in unstructured free-text fields of electronic health records supplement manually coded diagnoses. The median accuracy of diagnostic coding in discharge summaries is 80.3% [44], but the coding of comorbidities in problem lists is often incomplete [45–48]. The lack of completeness results in poor sensitivity of diagnostic coding, despite achieving high specificity [45, 49–52]. Therefore, the absence of a diagnostic code does not necessarily reflect the absence of the disease. Coding accuracy is further affected by changes in the coding systems used [47], lack of suitably granular codes [53], incomplete coding in single centres due to data fragmentation across multiple sites [54], and length of time registered in an EHR [55]. Supplementing diagnostic coding with unstructured fields can improve this sensitivity [2, 56, 57]. This increased sensitivity has important implications for the case-finding ability of studies using electronic health records.

Our workflow has several limitations. The NER model extracts entities as they appear in text and is not integrated with a linking process to standard ontology. Therefore, linking terms to an ontology is considered a downstream task. However, building a database of diagnostic entities as they appear in the clinical records can inform further development of linking strategies or vocabulary databases. Our model was trained and evaluated using clinical records from a single institution. The model's performance, if evaluated using external notes, is likely to be lower. However, rapid dataset annotation using low-code NLP tools means any institution can create custom NER models. Furthermore, annotations were performed by a single annotator. Thus, the registry represents the annotating characteristics of a single annotator. Multiple annotators may reduce this bias; however, annotators should be trained to follow annotation guidelines to ensure adequate inter-annotator agreement [57]. Lastly, all annotations were performed in English. Replication of the study findings with non-English free text would be beneficial.

We demonstrated a workflow using low-code NLP tools to produce an ophthalmic disease registry, with an accompanying ready-to-use tool to reproduce the registry in any institution. Our NER model displayed a moderate overall ability to extract ophthalmic diagnoses from free-text electronic clinical records. There is a further need for standard ophthalmic datasets for the evaluation of NER models and ready-to-use tools to encourage increased use of artificial intelligence for clinical NER tasks.

Declarations

Informed consent A waiver of informed consent to access deidentified data was approved by the Human Research Ethics Committee.

Conflict of interest The authors declare no competing interests.

Research involving human participants and/or animals All procedures performed in studies involving human participants were in accordance with the ethical standards of the Central Adelaide Local Health Network Human Research Ethics Committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

1. Wu Y, Jiang M, Xu J, Zhi D, Xu H (2017) Clinical named entity recognition using deep learning models. AMIA Annu Symp Proc 2017:1812–1819

2. Abhyankar S, Demner-Fushman D, Callaghan FM, McDonald CJ (2014) Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. J Am Med Inform Assoc 21:801–807. https://doi.org/10.1136/amiajnl-2013-001915

3. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG (2010) Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 17:507–513. https://doi.org/10.1136/jamia.2009.001560

4. McInnes BT, Stevenson M (2014) Determining the difficulty of word sense disambiguation. J Biomed Inform 47:83–90

5. Wu Y, Denny J, Rosenbloom S, Miller R, Giuse D, Song M, Xu H (2015) A preliminary study of clinical abbreviation disambiguation in real time. Appl Clin Inform 6:364–374

6. Moon S, Pakhomov S, Melton GB (2012) Automated disambiguation of acronyms and abbreviations in clinical texts: window and training size considerations. In: AMIA annual symposium proceedings. American Medical Informatics Association, p 1310

7. Ruch P, Baud R, Geissbühler A (2003) Using lexical disambiguation and named-entity recognition to improve spelling correction in the electronic patient record. Artif Intell Med 29:169–184

8. Edinger T, Cohen AM, Bedrick S, Ambert K, Hersh W (2012) Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC Medical Records Track. AMIA Annu Symp Proc 2012:180–188

9. Demner-Fushman D, Mork JG, Shooshan SE, Aronson AR (2010) UMLS content views appropriate for NLP processing of the biomedical literature vs. clinical text. J Biomed Inform 43:587–594. https://doi.org/10.1016/j.jbi.2010.02.005

10. Hanauer DA, Mei Q, Vydiswaran VGV, Singh K, Landis-Lewis Z, Weng C (2019) Complexities, variations, and errors of numbering within clinical notes: the potential impact on information extraction and cohort-identification. BMC Med Inform Decis Mak 19:75. https://doi.org/10.1186/s12911-019-0784-1

11. Skeppstedt M, Kvist M, Dalianis H (2012) Rule-based entity recognition and coverage of SNOMED CT in Swedish Clinical Text. LREC, pp 1250–1257

12. Chen L, Gu Y, Ji X, Lou C, Sun Z, Li H, Gao Y, Huang Y (2019) Clinical trial cohort selection based on multi-level rule-based natural language processing system. J Am Med Inform Assoc 26:1218–1226. https://doi.org/10.1093/jamia/ocz109

13. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, Xu H (2011) A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. J Am Med Inform Assoc 18:601–606. https://doi.org/10.1136/amiajnl-2011-000163

14. Yadav V, Bethard S (2018) A survey on recent advances in named entity recognition from deep learning models. Association for Computational Linguistics, Santa Fe, New Mexico, USA, pp 2145–2158

15. Fraile Navarro D, Ijaz K, Rezazadegan D, Rahimi-Ardabili H, Dras M, Coiera E, Berkovsky S (2023) Clinical named entity recognition and relation extraction using natural language processing of medical free text: a systematic review. Int J Med Inform 177:105122. https://doi.org/10.1016/j.ijmedinf.2023.105122

16. Hedderich DM, Keicher M, Wiestler B, Gruber MJ, Burwinkel H, Hinterwimmer F, Czempiel T, Spiro JE, Pinto dos Santos D, Heim D, Zimmer C, Rückert D, Kirschke JS, Navab N (2021) AI for doctors—a course to educate medical professionals in artificial intelligence for medical imaging. Healthcare 9:1278

17. Boillat T, Nawaz FA, Rivas H (2022) Readiness to embrace artificial intelligence among medical doctors and students: questionnaire-based study. JMIR Med Educ 8:e34973. https://doi.org/10.2196/34973

18. Chen M, Zhang B, Cai Z, Seery S, Gonzalez MJ, Ali NM, Ren R, Qiao Y, Xue P, Jiang Y (2022) Acceptance of clinical artificial intelligence among physicians and medical students: a systematic review with cross-sectional survey. Frontiers in Medicine 9:990604. https://doi.org/10.3389/fmed.2022.990604

19. Scheetz J, Rothschild P, McGuinness M, Hadoux X, Soyer HP, Janda M, Condon JJ, Oakden-Rayner L, Palmer LJ, Keel S (2021) A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep 11:1–10

20. GmbH E (2017-2023) Prodigy, ExplosionAI GmbH, Skalitzer Str. 100, 10997, Berlin, Germany

21. Macri C, Teoh I, Bacchi S, Sun M, Selva D, Casson R, Chan W (2022) Automated identification of clinical procedures in free-text electronic clinical records with a low-code named entity recognition workflow. Methods Inf Med 61:084–089. https://doi.org/10.1055/s-0042-1749358

22. Dalianis H (2018) Evaluation metrics and evaluation. Clinical text mining: secondary use of electronic patient records. Springer International Publishing, Cham, pp 45–53

23. Nejadgholi I, Fraser KC, De Bruijn B (2020) Extensive error analysis and a learning-based evaluation of medical entity recognition systems to approximate user experience. arXiv preprint arXiv:200605281

24. Singhal A, Buckley C, Mitra M (2017) Pivoted document length normalization ACM SIGIR Forum. ACM, New York, NY, USA, pp 176–184

25. Metabase (2023) Metabase, San Francisco, California, USA

26. Neves M, Ševa J (2021) An extensive review of tools for manual annotation of documents. Brief Bioinform 22:146–163. https://doi.org/10.1093/bib/bbz130

27. Lingren T, Deleger L, Molnar K, Zhai H, Meinzen-Derr J, Kaiser M, Stoutenborough L, Li Q, Solti I (2014) Evaluating the impact of pre-annotation on annotation speed and potential bias: natural language processing gold standard development for clinical named entity recognition in clinical trial announcements. J Am Med Inform Assoc 21:406–413. https://doi.org/10.1136/amiajnl-2013-001837

28. Palmer EL, Hassanpour S, Higgins J, Doherty JA, Onega T (2019) Building a tobacco user registry by extracting multiple smoking behaviors from clinical notes. BMC Med Inform Decis Mak 19:141. https://doi.org/10.1186/s12911-019-0863-3

29. Al-Haddad MA, Friedlin J, Kesterson J, Waters JA, Aguilar-Saavedra JR, Schmidt CM (2010) Natural language processing for the development of a clinical registry: a validation study in intraductal papillary mucinous neoplasms. HPB (Oxford) 12:688–695. https://doi.org/10.1111/j.1477-2574.2010.00235.x

30. Shah RF, Bini S, Vail T (2020) Data for registry and quality review can be retrospectively collected using natural language processing from unstructured charts of arthroplasty patients. Bone Joint J 102:99–104. https://doi.org/10.1302/0301-620x.102b7.Bjj-2019-1574.R1

31. Berman AN, Biery DW, Ginder C, Hulme OL, Marcusa D, Leiva O, Wu WY, Singh A, Divakaran S, Hainer J, Turchin A, Januzzi JL, Natarajan P, Cannon CP, Di Carli MF, Bhatt DL, Blankstein R (2020) Study of lipoprotein(a) and its impact on atherosclerotic cardiovascular disease: design and rationale of the Mass General Brigham Lp(a) Registry. Clin Cardiol 43:1209–1215. https://doi.org/10.1002/clc.23456

32. Oliwa T, Maron SB, Chase LM, Lomnicki S, Catenacci DVT, Furner B, Volchenboum SL (2019) Obtaining knowledge in pathology reports through a natural language processing approach with classification, named-entity recognition, and relation-extraction heuristics. JCO Clin Cancer Inform 3:1–8. https://doi.org/10.1200/cci.19.00008

33. Moqurrab SA, Ayub U, Anjum A, Asghar S, Srivastava G (2021) An accurate deep learning model for clinical entity recognition from clinical notes. IEEE J Biomed Health Inf 25:3804–3811. https://doi.org/10.1109/JBHI.2021.3099755

34. Si Y, Wang J, Xu H, Roberts K (2019) Enhancing clinical concept extraction with contextual embeddings. J Am Med Inform Assoc 26:1297–1304. https://doi.org/10.1093/jamia/ocz096

35. Kim Y, Riloff E, Hurdle JF (2015) A study of concept extraction across different types of clinical notes. AMIA Annu Symp Proc 2015:737–746

36. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A (2020) Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet 28:165–173. https://doi.org/10.1038/s41431-019-0508-0

37. Rath A, Salamon V, Peixoto S, Hivert V, Laville M, Segrestin B, Neugebauer EAM, Eikermann M, Bertele V, Garattini S, Wetterslev J, Banzi R, Jakobsen JC, Djurisic S, Kubiak C, Demotes-Mainard J, Gluud C (2017) A systematic literature review of evidence-based clinical practice for rare diseases: what are the perceived and real barriers for improving the evidence and how can they be overcome? Trials 18:556. https://doi.org/10.1186/s13063-017-2287-7

38. Aymé S, Bellet B, Rath A (2015) Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. Orphanet J Rare Dis 10:35. https://doi.org/10.1186/s13023-015-0251-8

39. Tisdale A, Cutillo CM, Nathan R, Russo P, Laraway B, Haendel M, Nowak D, Hasche C, Chan CH, Griese E, Dawkins H, Shukla O, Pearce DA, Rutter JL, Pariser AR (2021) The IDeaS initiative: pilot study to assess the impact of rare diseases on patients and healthcare systems. Orphanet J Rare Dis 16:429. https://doi.org/10.1186/s13023-021-02061-3

40. Sun AZ, Shu YH, Harrison TN, Hever A, Jacobsen SJ, O'Shaughnessy MM, Sim JJ (2020) Identifying patients with rare disease using electronic health record data: the Kaiser Permanente Southern California Membranous Nephropathy Cohort. Perm J 24. https://doi.org/10.7812/tpp/19.126

41. Garcelon N, Burgun A, Salomon R, Neuraz A (2020) Electronic health records for the diagnosis of rare diseases. Kidney Int 97:676–686. https://doi.org/10.1016/j.kint.2019.11.037

42. Lo Barco T, Kuchenbuch M, Garcelon N, Neuraz A, Nabbout R (2021) Improving early diagnosis of rare diseases using natural language processing in unstructured medical records: an illustration from Dravet syndrome. Orphanet J Rare Dis 16:309. https://doi.org/10.1186/s13023-021-01936-9

43. DeLozier S, Speltz P, Brito J, Tang LA, Wang J, Smith JC, Giuse D, Phillips E, Williams K, Strickland T, Davogustto G, Roden D, Denny JC (2021) Real-time clinical note monitoring to detect conditions for rapid follow-up: a case study of clinical trial enrollment in drug-induced torsades de pointes and Stevens-Johnson syndrome. J Am Med Inform Assoc 28:126–131. https://doi.org/10.1093/jamia/ocaa213

44. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, Faiz OD (2012) Systematic review of discharge coding accuracy. J Public Health (Oxf) 34:138–148. https://doi.org/10.1093/pubmed/fdr054

45. Bozic KJ, Bashyal RK, Anthony SG, Chiu V, Shulman B, Rubash HE (2013) Is administratively coded comorbidity and complication data in total joint arthroplasty valid? Clin Orthop Relat Res 471:201–205. https://doi.org/10.1007/s11999-012-2352-1

46. Nimmo A, Steenkamp R, Ravanan R, Taylor D (2021) Do routine hospital data accurately record comorbidity in advanced kidney disease populations? A record linkage cohort study. BMC Nephrol 22:95. https://doi.org/10.1186/s12882-021-02301-5

47. Nimptsch U (2016) Disease-specific trends of comorbidity coding and implications for risk adjustment in hospital administrative data. Health Serv Res 51:981–1001. https://doi.org/10.1111/1475-6773.12398

48. Wright A, McCoy AB, Hickman T-TT, Hilaire DS, Borbolla D, Bowes WA, Dixon WG, Dorr DA, Krall M, Malholtra S, Bates DW, Sittig DF (2015) Problem list completeness in electronic health records: a multi-site study and assessment of success factors. Int J Med Inform 84:784–790. https://doi.org/10.1016/j.ijmedinf.2015.06.011

49. Goff SL, Pekow PS, Markenson G, Knee A, Chasan-Taber L, Lindenauer PK (2012) Validity of using ICD-9-CM codes to identify selected categories of obstetric complications, procedures and co-morbidities. Paediatr Perinat Epidemiol 26:421–429. https://doi.org/10.1111/j.1365-3016.2012.01303.x

50. Higgins TL, Deshpande A, Zilberberg MD, Lindenauer PK, Imrey PB, Yu P-C, Haessler SD, Richter SS, Rothberg MB (2020) Assessment of the accuracy of using ICD-9 diagnosis codes to identify pneumonia etiology in patients hospitalized with pneumonia. JAMA Network Open 3:e207750–e207750. https://doi.org/10.1001/jamanetworkopen.2020.7750

51. Grams ME, Waikar SS, MacMahon B, Whelton S, Ballew SH, Coresh J (2014) Performance and limitations of administrative data in the identification of AKI. Clin J Am Soc Nephrol 9:682–689

52. Kern EF, Maney M, Miller DR, Tseng CL, Tiwari A, Rajan M, Aron D, Pogach L (2006) Failure of ICD-9-CM codes to identify patients with comorbid chronic kidney disease in diabetes. Health Serv Res 41:564–580

53. Navar AM (2019) Electronic health record data quality issues are not remedied by increasing granularity of diagnosis codes. JAMA Cardiology 4:465–465. https://doi.org/10.1001/jamacardio.2019.0830

54. Wei W-Q, Leibson CL, Ransom JE, Kho AN, Caraballo PJ, Chai HS, Yawn BP, Pacheco JA, Chute CG (2012) Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. J Am Med Inform Assoc 19:219–224

55. Wei W-Q, Leibson CL, Ransom JE, Kho AN, Chute CG (2013) The absence of longitudinal data limits the accuracy of high-throughput clinical phenotyping for identifying type 2 diabetes mellitus subjects. Int J Med Inform 82:239–247

56. Blecker S, Katz SD, Horwitz LI, Kuperman G, Park H, Gold A, Sontag D (2016) Comparison of approaches for heart failure case identification from electronic health record data. JAMA Cardiology 1:1014–1020. https://doi.org/10.1001/jamacardio.2016.3236

57. Chapman WW, Dowling JN, Hripcsak G (2008) Evaluation of training with an annotation schema for manual annotation of clinical conditions from emergency department reports. Int J Med Inform 77:107–113. https://doi.org/10.1016/j.ijmedinf.2007.01.002