

Distributed Knowledge Based Clinical Auto-Coding System

Rajvir Kaur

School of Computing, Engineering and Mathematics
Western Sydney University, Australia

18531738@student.westernsydney.edu.au

Abstract

Codification of free-text clinical narratives have long been recognised to be beneficial for secondary uses such as funding, insurance claim processing and research. In recent years, many researchers have studied the use of Natural Language Processing (NLP), related Machine Learning (ML) methods and techniques to resolve the problem of manual coding of clinical narratives. Most of the studies are focused on classification systems relevant to the U.S and there is a scarcity of studies relevant to Australian classification systems such as ICD-10-AM and ACHI. Therefore, we aim to develop a knowledge-based clinical auto-coding system, that utilise appropriate NLP and ML techniques to assign ICD-10-AM and ACHI codes to clinical records, while adhering to both local coding standards (Australian Coding Standard) and international guidelines that get updated and validated continuously.

1 Introduction

Documentation related to an episode of care of a patient, commonly referred to as a medical record, contains clinical findings, diagnoses, interventions, and medication details which are invaluable information for clinical decisions making. To carry out meaningful statistical analysis, these medical records are converted into a special set of codes which are called *Clinical codes* as per the clinical coding standards set by the World Health Organisation (WHO). The International Classification of Diseases (ICD) codes are a special set of alphanumeric codes, assigned to an episode of care of a patient, based on which reimbursement is done in some countries (Kaur and Ginige, 2018). Clinical codes are assigned by trained professionals, known as *clinical coders*, who have a sound knowledge of medical terminologies, clinical classification systems, and coding rules and guidelines. The current scenario of assigning clinical

codes is a manual process which is very expensive, time-consuming, and error-prone (Xie and Xing, 2018). The wrong assignment of codes leads to issues such as reviewing of whole process, financial losses, increased labour costs as well as delays in reimbursement process. The coded data is not only used by insurance companies for reimbursement purposes, but also by government agencies and policy makers to analyse healthcare systems, justify investments done in the healthcare industry and plan future investments based on these statistics (Kaur and Ginige, 2018).

With the transition from ICD-9 to ICD-10 in 1992, the number of codes increased from 3,882 codes to approximately 70,000, which further makes manual coding a non-trivial task (Subotin and Davis, 2014). On an average, a clinical coder codes 3 to 4 clinical records per hour, resulting in 15-42 records per day depending on the experience and efficiency of the human coder (Santos et al., 2008; Kaur and Ginige, 2018). The cost incurred in assigning clinical codes and their follow up corrections are estimated to be 25 billion dollars per year in the United States (Farkas and Szarvas, 2008; Xie and Xing, 2018). There are several reasons behind the wrong assignment of codes. First, assignment of ICD codes to patient's records is highly erroneous due to subjective nature of human perception (Arifoğlu et al., 2014). Second, manual process of assigning codes is a tedious task which leads to inability to locate critical and subtle findings due to fatigue. Third, in many cases, physicians or doctors often use abbreviations or synonyms, which causes ambiguity (Xie and Xing, 2018).

A study by (McKenzie and Walker, 2003), describes changes that have occurred in the coder workforce over the last eight years in terms of employment conditions, duties, resources, and access to and need for continuous education. Similarly,

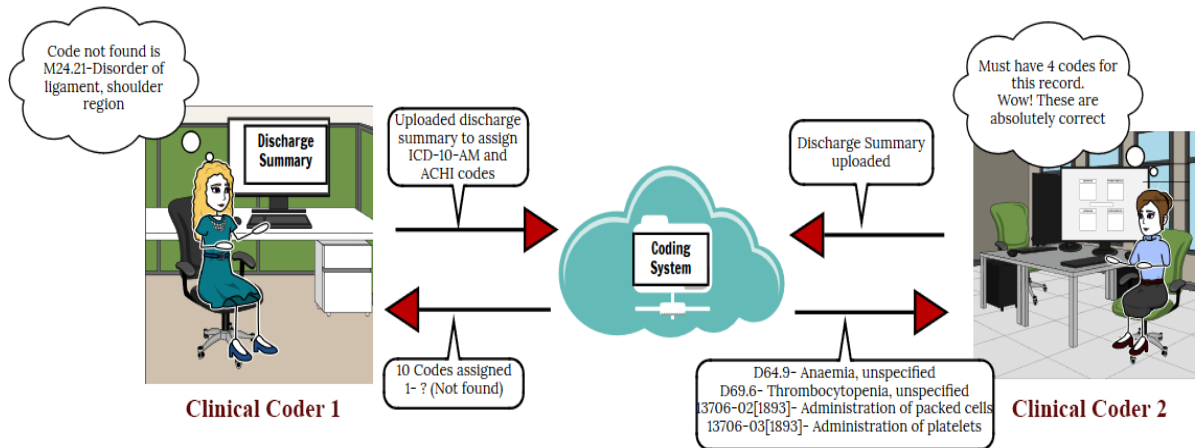


Figure 1: A distributed knowledge-based clinical auto-coding system

another study (Butler-Henderson, 2017), highlights major future challenges that health information management practitioners and academics will face with an ageing workforce, where more than 50% of the workforce is aged 45 years or older.

To reduce coding errors and cost, research is being conducted to develop methods for automated coding. Most of the research in auto-coding is focused on ICD-9-CM (Clinical Modification), ICD-10-CM, ICD-10-PCS (Procedure Coding System) which are US modifications. Very limited studies are focused on ICD-10-AM (Australian Modification) and Australian Classification of Health Intervention (ACHI). Hence, our research aims to develop a distributed knowledge-based clinical auto-coding system that would leverage on NLP and ML techniques, where a human coders will give their queries to the coding system and in revert the system will suggest a set of clinical codes. Figure 1 shows a possible scenario, how a distributed knowledge-based coding system will be used in practice.

2 Related Work

In early 19th century, a French statistician Jacques Bertillon, developed a classification system to record causes of death. Later in 1948, the WHO started maintaining the Bertillon classification and named it as International Statistical Classification of Disease, Injuries and Causes of Death (Cumerlato et al., 2010). Since then, roughly every ten years, this classification had been revised and in 1992, ICD-10 was approved. Twenty-six (26)

years after the introduction of ICD-10, the next generation of classification ICD-11 is released in May 2019 but not yet implemented (Kaur and Ginige, 2018). ICD-11 increases the complexity by introducing a new code structure, a new chapter on X-Extension Codes, dimensions of external causes (histopathology, consciousness, temporality, and etiology), and a new chapters on sleep-awake disorder, conditions related to sexual health, and traditional medicine conditions (Organisation, 2016; Hargreaves and Njeru, 2014; Reed et al., 2016).

In previous research related to clinical narrative analysis, different methods and techniques ranging from pattern matching to deep learning approaches are applied to categorise clinical narratives into different categories (Mujtaba et al., 2019). Several researchers across the globe have employed text classification to categorise clinical narratives into various categories using machine learning approaches including supervised (Hastie et al., 2009), unsupervised (Ko and Seo, 2000), semi-supervised (Zhu and Goldberg, 2009), ontology-based (Hotho et al., 2002), rule-based (Deng et al., 2015), transfer (Pan and Yang, 2010), and multi-view learning (Amini et al., 2009).

(Cai et al., 2016) reviewed the fundamentals of NLP and describe various techniques such as pattern matching, linguistic approach, statistical and machine learning approaches that constitute NLP in radiology, along with some key applications. (Larkey and Croft, 1995) studied three different classifiers namely: k-nearest neighbor, rel-

evance feedback and Bayesian independence classifiers for assigning ICD-9 codes to dictated inpatient discharge summaries. The study found that a combination of different classifiers produced better results than any single type of classifier. (Farkas and Szarvas, 2008) proposed a rule-based ICD-9-CM coding system for radiology reports and achieved good classification performances on a limited number of ICD-9-CM codes (45 in total). Similarly, (Goldstein et al., 2007; Pestian et al., 2007b; Crammer et al., 2007) also proposed automated system for assigning ICD-9-CM codes to free text radiology reports.

(Koopman et al., 2015) proposed a system for automatic ICD-10 classification of cancer from free-text death certificates. The classifiers were deployed in a two-level cascaded architecture, where the first level identifies the presence of cancer (i.e., binary form cancer/no cancer), and the second level identifies the type of cancer. However, all ICD-10 codes were truncated into three character level.

All the above mentioned research studies are based on some type of deep learning, machine learning or statistical approach, where the information contained in the training data is distillate into mathematical models, which can be successfully employed for assigning ICD codes (Chiaravalloti et al., 2014). One of the main flaws in these approaches is that training data is annotated by human coders. Thus, there is a possibility of inaccurate ICD codes. Therefore, if clinical records labelled with incorrect ICD codes are given as an input to an algorithm, it is likely that the model will also provide incorrect predictions.

2.1 Standard Pipeline for Clinical Text Classification

Various research studies have used different methods and techniques to handle and process clinical text, but the standard pipeline is utilised in some shape or form. This section details the steps in the standard pipeline in machine learning, as it is required for the auto-coding.

2.1.1 Types of clinical record

Clinical text classification techniques have been employed on different types of clinical records such as surgical reports (Stocker et al., 2014; Raja et al., 2012), radiology reports (Mendona et al., 2005), autopsy reports (Mujtaba et al., 2018), death certificates (Koopman et al., 2015), clinical

narratives (Meystre and Haug, 2006; Friedlin and McDonald, 2008), progress notes (Frost et al., 2005), laboratory reports (Friedlin and McDonald, 2008; Liu et al., 2012), admission notes and patient summaries (Jensen et al., 2012), pathology reports (Imler et al., 2013), and unstructured electronic text (Portet et al., 2009). In this research, we aim to primarily use clinical discharge summaries as the input text data.

2.1.2 Datasets available

The data sources used in various research studies can be categorised into two types: homogeneous sources and heterogeneous sources, which can further be divided into three subtypes: binary class, multi-class single labeled, multi-class multi-labeled datasets (Mujtaba et al., 2019). There are few datasets that are publicly available such as PhysioNet¹, i2b2 NLP dataset², and OHSUMED³. In this research, we aim to use both publicly available and data acquired from hospitals.

2.1.3 Preprocessing

Preprocessing is done to remove meaningless information from the dataset as the clinical narratives may contain high level of noise, sparsity, misspelled words, grammatical errors (Nguyen and Patrick, 2016; Mujtaba et al., 2019). Different preprocessing techniques are applied in research studies including sentence splitting, tokenisation, spell error detection and correction, stemming and lemmatisation, normalisation (Manning et al., 2008), removal of stop words, removal of punctuation or special symbols, abbreviation expansion, chunking, named entity recognition (Bird et al., 2009), negation detection (Chapman et al., 2001).

2.1.4 Feature Engineering

Feature engineering is the combination of feature extraction, feature representation, and feature selection (Mujtaba et al., 2019). Feature extraction is the process of extracting useful features which includes Bag of Words (BoW), n-gram, Word2Vec, and GloVe. Once features are extracted, next step is to represent in numeric form to feature vectors using either binary representation, term frequency (tf), term frequency with inverse document frequency (tf-idf), or normalised tf-idf.

¹<https://physionet.org/mimic2/>

²<https://www.i2b2.org/NLP/DataSets/>

³<http://davis.wpi.edu/xmdv/datasets/ohsumed.html>

2.1.5 Classification

For classification, various research studies have used classifiers such as Support Vector Machine (SVM) (Cortes and Vapnik, 1995), k-Nearest Neighbor (kNN) (Altman, 1992), Convolutional Neural Network (CNN) (Karimi et al., 2017), Recurrent Neural Network (RNN), Long short-term memory (LSTM)(Luo, 2017), and Gated Recurrent Unit (GRU) (Jagannatha and Yu, 2016).

2.1.6 Evaluation Metrics

The performance of clinical text classification models can be measured using standard evaluation metrics which include precision, recall, F-measure (or F-score), accuracy, precision (micro and macro-average), recall (micro and macro-average), F-measure (micro and macro-average), and area under the curve (AUC). These metrics can be computed by using values of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) in the standard confusion matrix (Mujtaba et al., 2019).

3 Experimental Framework

3.1 Data collection and ethics approval

This research has ethics approval from Western Sydney University Human Research Ethics Committee (HREC) under reference No: H12628 to use 1,200 clinical records. The ethics approval is valid for the next four years until 11th April, 2023. In addition, we also have access to publicly available dataset such as MIMIC-III and Computational Medicine Center (CMC) (Pestian et al., 2007a). Apart from this, more clinical records from acute or sub-acute hospitals will also be collected.

3.2 Proposed Research

Within the broader scope of this proposal, the work will be focused on the research questions given below:

How to optimise the use of computerised algorithms to assign ICD-10-AM and ACHI codes to clinical records, while adhering to local coding standard (for example, Australian Coding Standard (ACS)) and international guidelines, leveraging on a distributed knowledge-base?

To address main research question, the following sub-research questions will be investigated:

Why do certain algorithms perform differently with similar dataset?

The *No free lunch theorem* (Wolpert, 1996) states that there is no such algorithm that is universally best for every problem. If one algorithm does really good for a given dataset, it may not do really well for other dataset. For example, one cannot say that SVM always does better prediction than Naïve Bayes or Decision Tree all the times. The intention of ML or statistical learning research is not to find the universally best algorithm, but the reason is that most of the algorithms work on the sample data and then make predictions or inference out of that. We cannot make proper truthful prediction just by working on a sample data. In fact, the results are all probabilistic in nature, not 100% true or certain. The study (Kaur and Ginige, 2018), performed comparative analysis on different approaches such as pattern matching, rule-based, ML, and hybrid. Each of the above mentioned methods and techniques performed differently in every case, but there was no explanations given behind the performance of each algorithm. Moreover, this study did not used ACS rules while assigning ICD-10-AM and ACHI codes.

There are few reasons that may have effected the algorithms performance used for codification of ICD-10-AM and ACHI codes in the previous study (Kaur and Ginige, 2018). Firstly, domain knowledge is very essential before assigning codes. In Australia, coding standards are used for clinical coding purpose to provide consistency of data collection, and support secondary classifications based on ICD such as the Australian Refined Diagnosis Related Groups (AR-DRGs). Therefore, during ICD-10-AM and ACHI code assignment, ACS rules are considered. If these ACS rules are not considered, then there is a possibility of wrong assignment of codes. Secondly, the study (Kaur and Ginige, 2018) had very limited number of medical records due to which the algorithms were unable to learn and predict correct codes properly. A similar study (Kaur and Ginige, 2019) done by the same set of authors using the same dataset describes that the dataset contains 420 unique labels, out of which 221 labels appeared only once in the whole dataset, 77 labels appeared twice, and only 24 labels appeared more than 15 times. Therefore, it lowers the learning rate of the algorithms.

To overcome the above stated problems, we

will make use of ACS in conjunction in ICD-10-AM and ACHI codes, and use large-scale data so that the algorithms can learn properly and make correct predictions. In order to process raw data, feature engineering will be carried out to transform the raw data into feature vectors. Moreover, in NLP, word embeddings has the ability to capture high-level semantic and syntactic properties of text. A study by (Henriksson et al., 2015) leverages word embeddings to identify adverse drug events from clinical notes and shows that using word embeddings can improve the predictive performance of machine learning methods. Therefore, in our research, we will explore semantic and syntactic properties of text to improve the performance of algorithms which give different performance on the same dataset.

How to assign ICD codes before referring to local and international standards and guidelines?

In the U.S, the Centers for Medicare and Medicaid Services (CMS) and the National Center for Health Statistics (NCHS), provide the guidelines for coding and reporting using the ICD-10-CM. These guidelines are a set of rules that have been approved by the four organisations: American Hospital Association (AHA), the American Health Information Association (AHIMA), CMS, and NCHS (for Health Statistics). Similarly, in Australia, the clinical coding standards *i.e.*, ACS rules are designed to be used in conjunction with ICD-10-AM and ACHI and are applicable to all public and private hospitals in Australia (for Classification Development, 2017). The clinical codes are not only assigned based on the information provided on the front sheet or the discharge summary but a complete analysis is performed by following the guidelines given in the ACS.

Since the introduction of ICD-10 in 1992, many countries have modified the WHO’s ICD-10 classification system into their country specific reporting purpose. For example, ICD-10-CA (Canadian Modification) and ICD-10-GM (German Modification). There are few major difference between the US and Australian classification systems. Firstly, there are few additional ICD-10-AM codes that are more specific (approximately 4,915 codes) that are coded only in Australia and 15 other countries including Ireland, and Saudi Arabia that use Australian classification system as their national classification system. For exam-

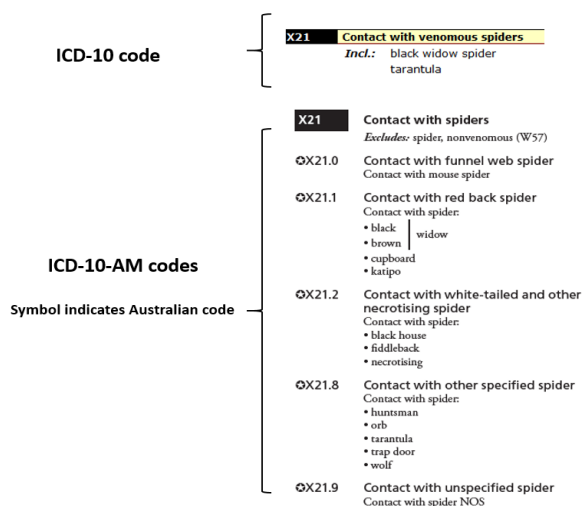


Figure 2: Difference between ICD-10 and ICD-10-AM codes.

ple, in the U.S, *contact with venomous spiders* is coded as X21, whereas in Australia, it is more specific by adding fourth character level as shown in Figure 2. There are 12% ICD-10-AM specific codes that do not exist in ICD-10-CM, ICD-9-CM or any other classification system. Secondly, countries that have developed their own national classification system use different coding practices. For example, in the U.S, Pulmonary oedema is coded as J81, whereas in Australia, to assign code for Pulmonary oedema, there is ACS rule 0920 which says, “When *acute pulmonary oedema* is documented without further qualification about the underlying cause, assign *I50.1 Left ventricular failure*”. Therefore, in our research, we will find methods and techniques to represent the coding standards and guidelines in a computerised format before assigning ICD codes. In addition, we will also explore mechanisms to manage the evolving nature of coding standards.

How to pre-process heterogeneous dataset?

Collecting data in health-care domain is a challenge in itself. Though, there are few publicly available repositories, there are certain issues to be resolved before using these in our research. For example, MIMIC dataset contains de-identified health data based on ICD-9 codes and Current Procedural Terminology (CPT) codes. As our research is focused on assigning ICD-10-AM and ACHI codes to clinical records, there is a need of mapping between ICD-9 to ICD-10 and vice-versa and ICD-10-CM to ICD-10-AM.

There are some existing look-up, translators, or mapping tools, which will translate ICD-9 codes into ICD-10 codes and vice versa (Butler, 2007). Therefore, we will explore and use the existing mapping tools to convert ICD-9 to ICD-10 codes, ICD-10 to ICD-10-AM codes or another classification system in order to train the model that is not annotated using ICD-10-AM and ACHI codes.

What sort of a distributed knowledge-based system would support the assigning clinical codes?

The majority of studies have used ML, hybrid, and deep learning approaches for clinical text classification. There are two main challenges that one has to face while doing research in health-care domain. First, to train the model when data is scarce. The ML based algorithms for classification and automated ICD code assignment are characterised by many limitations. For example, knowledge acquisition bottleneck, in which ML algorithms require a large number of annotated data for constructing an accurate classification model. Therefore, many believe that the quality of ML based algorithms highly depended on data rather than algorithms (Mujtaba et al., 2019). Even after a great efforts, researchers are able to collect millions of data, there is still a possibility that the occurrence of some diseases and interventions will not be enough to train the model properly and give correct codes. However, when data is insufficient, transfer learning or fine tuning are other possible options to look into (Singh, 2018). Secondly, it is difficult and expensive to assign ground truth codes (or label) to the clinical records. Although, the above mentioned approaches are capable of providing good results, but these approaches require annotated data in order to train the model. The labelling process requires human expert to assign labels (or ICD codes) to each clinical record. For example, the study (Kaur and Ginige, 2018) contains 190 de-identified discharge summaries belonging to diseases and interventions of respiratory and digestive system. The discharge summaries were in the hand written form, which were later converted into digital form and assigned ground truth codes with the help of a human expert. Thus, a considerable amount of effort was exerted in preparing the training data.

Therefore, in our research we aim to develop a distributed knowledge-base system where humans

(clinical coders) and machines can work together to overcome the above mentioned challenges. If machine is unable to predict the correct ICD code for a given disease or intervention then humans input will be considered. Moreover, the human coder can also verify the codes assigned by machine.

3.3 Baseline Methods

There are three main approaches for automated ICD codes assignment: (1) machine learning; (2) hybrid (combining machine learning and rule-base); and (3) deep learning. Deep learning models have demonstrated successful results in many NLP tasks such as language translation (Zhang and Zong, 2015), image captioning (LeCun et al., 2015) and sentiment analysis (Socher et al., 2013). We will work on different ML and deep learning models including LSTM, CNN-RNN, and GRU. Pre-processing will be done using standard pipeline and convert the assigned labels based on Australian classification system using existing mapping tools. Feature extraction will be done using non-sequential and sequential features followed by training and testing of the model using baseline models and deep learning models.

4 Conclusion

In this research proposal, we aim to develop a knowledge-based clinical auto-coding system that uses computerised algorithms to assign ICD-10-AM, ACHI, ICD-11, and ICHI codes to an episode of care of a patient while adhering coding guidelines. Further, we will explore how ML models can be trained with limited dataset, mapping between different classification systems, and avoiding labelling efforts.

Acknowledgments

I would like to thank my supervisors, Dr. Jeewani Anupama Ginige, Dr. Oliver Obst and Vera Dimitropoulos for their feedback that greatly improved the paper. This research is supported by Western Sydney University Postgraduate Research Scholarship.

References

N. S. Altman. 1992. [An introduction to kernel and nearest-neighbor nonparametric regression](#). *The American Statistician*, 46(3):175–185.

- M.R. Amini, N. Usunier, and C. Goutte. 2009. [Learning from multiple partially observed views - an application to multilingual text categorization](#). pages 28–36, Vancouver, BC. Conference of 23rd Annual Conference on Neural Information Processing Systems, NIPS 2009 ; Conference Date: 7 December 2009 - 10 December 2009.
- Damla Arifoğlu, Onur Deniz, Kemal Aleçakır, and Meltem Yöndem. 2014. Codemagic: Semi-automatic assignment of ICD-10-AM codes to patient records. In *Information Sciences and Systems 2014*, pages 259–268. Springer International Publishing.
- Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: analyzing text with the natural language toolkit*. ”O’Reilly Media, Inc.”.
- Rhonda R Butler. 2007. ICD-10 general equivalence mappings: Bridging the translation gap from ICD-9. *Journal of American Health Information Management Association*, 78(9):84–86.
- Kerryn Butler-Henderson. 2017. Health information management 2025. what is required to create a sustainable profession in the face of digital transformation?
- Tianrun Cai, Andreas A. Giannopoulos, Sheng Yu, Tatiana Kelil, Beth Ripley, Kanako K. Kumamaru, Frank J. Rybicki, and Dimitrios Mitsouras. 2016. [Natural language processing technologies in radiology research and clinical applications](#). *RadioGraphics*, 36(1):176–191. PMID: 26761536.
- Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001. [A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries](#). *Journal of Biomedical Informatics*, 34(5):301 – 310.
- Maria Teresa Chiaravalloti, Roberto Guarasci, Vincenzo Lagani, Erika Pasceri, and Roberto Trunfio. 2014. [A coding support system for the icd-9-cm standard](#). In *2014 IEEE International Conference on Healthcare Informatics*, pages 71–78.
- Australian Consortium for Classification Development. 2017. *Australian Coding Standards for ICD-10-AM and ACHI*. Independent Hospital Pricing Authority.
- Corinna Cortes and Vladimir Vapnik. 1995. [Support-Vector Networks](#). *Machine Learning*, 20(3):273–297.
- Koby Crammer, Mark Dredze, Kuzman Ganchev, Partha Pratim Talukdar, and Steven Carroll. 2007. [Automatic code assignment to medical text](#). In *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, BioNLP ’07, pages 129–136, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Megan Cumerlato, Lindy Best, Belinda Saad, and N.S.W.) National Centre for Classification in Health (Sydney. 2010. *Fundamentals of morbidity coding using ICD-10-AM, ACHI, and ACS seventh edition*. Sydney : National Centre for Classification in Health. ”June 2010”–T.p.
- Y. Deng, M.J. Groll, and K. Denecke. 2015. [Rule-based cervical spine defect classification using medical narratives](#). *Studies in Health Technology and Informatics*, 216:1038. Conference of 15th World Congress on Health and Biomedical Informatics, MEDINFO 2015 ; Conference Date: 19 August 2015-23 August 2015.
- Richárd Farkas and György Szarvas. 2008. [Automatic construction of rule-based ICD-9-CM coding systems](#). *BMC Bioinformatics*, 9(3):S10.
- F. Jeff Friedlin and Clement J. McDonald. 2008. [A Software Tool for Removing Patient Identifying Information from Clinical Documents](#). *Journal of the American Medical Informatics Association*, 15(5):601–610.
- H. Robert Frost, Dean F. Sittig, Victor J. Stevens, and Brian Hazlehurst. 2005. [MediClass: A System for Detecting and Classifying Encounter-based Clinical Events in Any Electronic Medical Record](#). *Journal of the American Medical Informatics Association*, 12(5):517–529.
- Ira Goldstein, Anna Arzumtsyan, and Özlem Uzuner. 2007. Three approaches to automatic assignment of icd-9-cm codes to radiology reports. In *AMIA Annual Symposium Proceedings*, volume 2007, page 279. American Medical Informatics Association.
- Jenny Hargreaves and Jodee Njeru. 2014. ICD-11: A dynamic classification for the information age.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *Overview of Supervised Learning*, pages 9–41. Springer New York, New York, NY.
- National Center for Health Statistics. [ICD-10-CM Official Guidelines for Coding and Reporting](#).
- Aron Henriksson, Maria Kvist, Hercules Dalianis, and Martin Duneld. 2015. [Identifying adverse drug event information in clinical notes with distributional semantic representations of context](#). *Journal of Biomedical Informatics*, 57:333 – 349.
- Andreas Hotho, Alexander Maedche, and Steffen Staab. 2002. Ontology-based text document clustering. *KI*, 16(4):48–54.
- Timothy D. Imler, Justin Morea, Charles Kahi, and Thomas F. Imperiale. 2013. [Natural language processing accurately categorizes findings from colonoscopy and pathology reports](#). *Clinical Gastroenterology and Hepatology*, 11(6):689 – 694.

- Abhyuday N Jagannatha and Hong Yu. 2016. Bidirectional RNN for medical event detection in electronic health records. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, volume 2016, page 473. NIH Public Access.
- Peter B Jensen, Lars J Jensen, and Søren Brunak. 2012. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, 13(6):395.
- Sarvnaz Karimi, Xiang Dai, Hamedh Hassanzadeh, and Anthony Nguyen. 2017. Automatic diagnosis coding of radiology reports: A comparison of deep learning and conventional classification methods. In *BioNLP 2017*, pages 328–332. Association for Computational Linguistics.
- Rajvir Kaur and Jeewani Anupama Ginige. 2018. Comparative Analysis of Algorithmic Approaches for Auto-Coding with ICD-10-AM and ACHI. *Studies in Health Technology and Informatics*, 252:73–79.
- Rajvir Kaur and Jeewani Anupama Ginige. 2019. Analysing effectiveness of multi-label classification in clinical coding. In *Proceedings of the Australasian Computer Science Week Multiconference, ACSW 2019*, pages 24:1–24:9, New York, NY, USA. ACM.
- Youngjoong Ko and Jungyun Seo. 2000. Automatic text categorization by unsupervised learning. In *Proceedings of the 18th Conference on Computational Linguistics - Volume 1, COLING '00*, pages 453–459, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Bevan Koopman, Guido Zuccon, Anthony Nguyen, Anton Bergheim, and Narelle Grayson. 2015. Automatic ICD-10 classification of cancers from free-text death certificates. *International Journal of Medical Informatics*, 84(11):956 – 965.
- Leah S Larkey and W Bruce Croft. 1995. Automatic assignment of ICD-9 codes to discharge summaries. Technical report, Technical report, University of Massachusetts at Amherst, Amherst, MA.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature*, 521(7553):436.
- Hongfang Liu, Kavishwar B Waghlikar, Kathy L MacLaughlin, Michael R Henry, Robert A Greenes, Ronald A Hankey, and Rajeev Chaudhry. 2012. Clinical decision support with automated text processing for cervical cancer screening. *Journal of the American Medical Informatics Association*, 19(5):833–839.
- Yuan Luo. 2017. Recurrent neural networks for classifying relations in clinical notes. *Journal of Biomedical Informatics*, 72:85 – 95.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
- Kirsten McKenzie and Sue M Walker. 2003. *The Australian coder workforce 2002: a report of the National Clinical Coder Survey*. National Centre for Classification in Health.
- Eneida A. Mendona, Janet Haas, Lyudmila Shagina, Elaine Larson, and Carol Friedman. 2005. Extracting information on pneumonia in infants using natural language processing of radiology reports. *Journal of Biomedical Informatics*, 38(4):314 – 321.
- Stphane Meystre and Peter J. Haug. 2006. Natural language processing to extract medical problems from electronic clinical documents: Performance evaluation. *Journal of Biomedical Informatics*, 39(6):589 – 599.
- Ghulam Mujtaba, Liyana Shuib, Norisma Idris, Wai Lam Hoo, Ram Gopal Raj, Kamran Khowaja, Khairunisa Shaikh, and Henry Friday Nweke. 2019. Clinical text classification research trends: Systematic literature review and open issues. *Expert Systems with Applications*, 116:494 – 520.
- Ghulam Mujtaba, Liyana Shuib, Ram Gopal Raj, Retnagowri Rajandram, and Khairunisa Shaikh. 2018. Prediction of cause of death from forensic autopsy reports using text classification techniques: A comparative study. *Journal of Forensic and Legal Medicine*, 57:41 – 50. Thematic section: Big data Guest editor: Thomas Lefvre Thematic section: Health issues in police custody Guest editors: Patrick Chariot and Steffen Heide.
- Hoang Nguyen and Jon Patrick. 2016. Text mining in clinical domain: Dealing with noise. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pages 549–558, New York, NY, USA. ACM.
- World Health Organisation. 2016. *ICD-11 Revision Conference Report*.
- Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.
- John P. Pestian, Chris Brew, Pawel Matykiewicz, DJ Hovermale, Neil Johnson, K. Bretonnel Cohen, and Wlodzislaw Duch. 2007a. A shared task involving multi-label classification of clinical free text. In *Biological, translational, and clinical language processing*, pages 97–104. Association for Computational Linguistics.
- John P. Pestian, Christopher Brew, Pawel Matykiewicz, D. J. Hovermale, Neil Johnson, K. Bretonnel Cohen, and W Duch. 2007b. A shared task involving

- multi-label classification of clinical free text. In *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, BioNLP '07, pages 97–104, Stroudsburg, PA, USA. Association for Computational Linguistics.
- François Portet, Ehud Reiter, Albert Gatt, Jim Hunter, Somayajulu Sripada, Yvonne Freer, and Cindy Sykes. 2009. [Automatic generation of textual summaries from neonatal intensive care data](#). *Artificial Intelligence*, 173(7-8):789–816. AvImpFact=2.566 estim. in 2012.
- Ali S. Raja, Ivan K. Ip, Luciano M. Prevedello, Aaron D. Sodickson, Cameron Farkas, Richard D. Zane, Richard Hanson, Samuel Z. Goldhaber, Ritu R. Gill, and Ramin Khorasani. 2012. [Effect of computerized clinical decision support on the use and yield of ct pulmonary angiography in the emergency department](#). *Radiology*, 262(2):468–474. PMID: 22187633.
- Geoffrey M. Reed, Jack Drescher, Richard B. Krueger, Elham Atalla, Susan D. Cochran, Michael B. First, Peggy T. Cohen-Kettenis, Ivn Arango-de Montis, Sharon J. Parish, Sara Cottler, Peer Briken, and Shekhar Saxena. 2016. [Disorders related to sexuality and gender identity in the icd-11: revising the icd-10 classification based on current scientific evidence, best clinical practices, and human rights considerations](#). *World Psychiatry*, 15(3):205–221.
- Suong Santos, Gregory Murphy, Kathryn Baxter, and Kerin M Robinson. 2008. [Organisational factors affecting the quality of hospital clinical coding](#). *Health Information Management Journal*, 37(1):25–37.
- Sonit Singh. 2018. [Pushing the limits of radiology with joint modeling of visual and textual information](#). In *Proceedings of ACL 2018, Student Research Workshop*, pages 28–36. Association for Computational Linguistics.
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013. [Recursive deep models for semantic compositionality over a sentiment treebank](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642. Association for Computational Linguistics.
- Christof Stocker, Leopold-Michael Marzi, Christian Matula, Johannes Schantl, Gottfried Prohaska, Aberto Brabenetz, and Andreas Holzinger. 2014. [Enhancing patient safety through human-computer information retrieval on the example of german-speaking surgical reports](#). In *2014 25th International Workshop on Database and Expert Systems Applications*, pages 216–220.
- Michael Subotin and Anthony Davis. 2014. [A system for predicting icd-10-pcs codes from electronic health records](#). In *Proceedings of BioNLP 2014*, pages 59–67, Baltimore, Maryland. Association for Computational Linguistics.
- David H. Wolpert. 1996. [The lack of a priori distinctions between learning algorithms](#). *Neural Computation*, 8(7):1341–1390.
- Pengtao Xie and Eric Xing. 2018. [A neural architecture for automated icd coding](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1066–1076. Association for Computational Linguistics.
- J. Zhang and C. Zong. 2015. [Deep neural networks in machine translation: An overview](#). *IEEE Intelligent Systems*, 30(5):16–25.
- Xiaojin Zhu and Andrew B Goldberg. 2009. [Introduction to semi-supervised learning](#). *Synthesis lectures on artificial intelligence and machine learning*, 3(1):1–130.