Medical Subdomain Classification of Clinical Notes Using a Machine Learning-Based Natural Language Processing Approach

Medical Subdomain Classification of Clinical Notes
Using a Machine Learning-Based Natural Language Processing Approach

By

Wei-Hung Weng, M.D.

Thesis submitted to the Harvard Medical School

In Partial Fulfillment of the Requirements for the Degree of

Master of Medical Sciences in Biomedical Informatics

at the

Harvard Medical School

Boston, Massachusetts

May 2017

Signature of Author……………………………………………….. ………………
Harvard Medical School Date

Certified by……………………………………………………….. ………………
Director, MMSc in Biomedical Informatics Date
Alexa T. McCray, Ph.D.
Professor of Medicine
Harvard Medical School

Thesis Committee

By signing this, I am confirming that I have reviewed this thesis. It represents work done by the author under my supervision and guidance.

Accepted by……………………………………………………….... ………………
Thesis Committee Chair Date
Henry C. Chueh, M.D., M.S.
Assistant Professor of Medicine
Massachusetts General Hospital

Accepted by…………………………………………………………. ………………
Thesis Committee Member Date
Alexa T. McCray, Ph.D.
Professor of Medicine
Harvard Medical School

Accepted by…………………………………………………………. ………………
Thesis Committee Member Date
Peter Szolovits, Ph.D.
Professor of Computer Science and Engineering
Massachusetts Institute of Technology

Medical Subdomain Classification of Clinical Notes Using a Machine Learning-Based Natural Language Processing Approach

Wei-Hung Weng, MD[1,2], Kavishwar B.Wagholikar, MBBS, PhD[2,3], Alexa T. McCray, PhD[1], Peter Szolovits, PhD[4], Henry C. Chueh, MD, MS[2,3]

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA;

[2]Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA;

[3]Department of Medicine, Massachusetts General Hospital, Boston, MA;

[4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Corresponding author: Wei-Hung Weng, MD. 50 Staniford Street, Suite 750, Boston, MA 02114. ckbjimmy@gmail.com. (857) 400-4997.

Keywords: Medical Decision Making, Computer-assisted; Natural Language Processing; Unified Medical Language System; Machine Learning

Word count - Abstract: 249

Word count - Main text: 3696

Number of tables: 4

Number of figures: 3

Number of references: 30

Number of supplementary tables: 3

Number of supplementary figures: 2

ABSTRACT

OBJECTIVE:

The medical subdomain of a clinical note, such as cardiology or neurology, is useful content-derived metadata for developing machine learning downstream applications. To classify the medical subdomain of a note accurately, we have constructed a machine learning-based natural language processing (NLP) pipeline and developed medical subdomain classifiers based on the content of the note.

MATERIALS AND METHODS:

We constructed the pipeline using the clinical NLP system, clinical Text Analysis and Knowledge Extraction System (cTAKES), the UMLS Metathesaurus, Semantic Network, and learning algorithms to extract features from two datasets — clinical notes from Integrating Data for Analysis, Anonymization, and Sharing (iDASH) data repository (n = 431) and Massachusetts General Hospital (MGH) (n = 91,237), and built medical subdomain classifiers with different combinations of clinical feature representations and learning algorithms. We evaluated the performance of classifiers and their portability across the two datasets.

RESULTS:

The linear support vector machine-trained medical subdomain classifier using hybrid bag-of-words and clinically relevant UMLS concepts as the feature representation, with term frequency-inverse document frequency (tf-idf)-weighting, outperformed other classifiers on iDASH and MGH datasets with F1 scores of 0.932 and 0.934, and areas under curve (AUC) of 0.957 and 0.964, respectively. We trained classifiers on one dataset, applied to the other dataset and yielded the threshold of F1 score of 0.7 in classifiers for half of medical subdomains.

CONCLUSION:

Our study shows that a supervised learning-based NLP approach is useful to develop medical subdomain classifiers. Portable classifiers may also be used across datasets from different institutions.

BACKGROUND AND SIGNIFICANCE

Automated document classification is an effective method that can categorize the documents into predefined document-level thematic labels.[1] Clinical notes, in which the medical reports are mainly written in natural language, have been regarded as a powerful resource to solve different clinical questions by providing detailed patient conditions, the thinking process of clinical reasoning, and clinical inference, which usually cannot be obtained from the other components of the electronic health record (EHR) system (e.g., claims data or laboratory examinations). Automated document classification is generally helpful in further processing clinical documents to extract these kinds of data. As such, the massive generation of clinical notes and rapidly increasing adoption of EHR systems has caused automated document classification to become an important research field of clinical predictive analytics, to help leverage the utility of narrative clinical notes.[2]

The medical subdomain, such as cardiology, gastroenterology and neurology, may be useful to enhance the effectiveness of clinical predictive analytics by considering specialty-associated conditions.[3] Knowing the medical subdomain helps with subsequent steps in data and knowledge extraction. Training on specialist reports and applying the subdomain models on notes written by generalists, such as general practitioners and internists, will also help identify the major problems of the patient that are being described. This can be useful not only in studying the practice and validity of clinical referral patterns, but also in helping to focus attention on the most pressing medical problem subdomain of the patient.

In the past, automated document classification has often been performed via rule-based knowledge engineering, by manually implementing a set of expert intelligence rules.[1] More recently, machine learning and natural language processing (NLP) techniques have been utilized to discover new clinical knowledge and develop clinical decision support systems from clinical documents.[4-9] Recently, several methods have been reported to classify MEDLINE documents, for example by using a hybrid word and phrase representation with a support vector machine (SVM) learning algorithm,[10] or adopting the Medical Subject Headings (MeSH) ontology as a feature representation with a maximum entropy algorithm to classify MEDLINE documents.[11] In order to classify clinical documents, Wilcox et al. used the Medical Language Extraction and Encoding System (MedLEE) with Unified Medical Language System (UMLS) Metathesaurus to identify medical concepts and classify chest radiograph reports into six clinical conditions.[12,13] D'Avolio et al. developed a clinical document processing system, automated retrieval console (ARC), to identify the presence of cancer in three sets of image and pathology reports.[14] However, integrating the medical subdomain information to classify real-world unstructured clinical notes using a learning-based NLP approach has not been investigated.

Development of machine learning classifiers for categorizing clinical notes, which have not been annotated or tagged, maximize the utility of the notes for clinical downstream applications in the medical specialty level. For example, using the medical subdomain classifier may help understand the language structure in the specific medical specialties, or more clinically, redirect patients with unsolved problems to the correct medical specialty for the appropriate management.

We developed a supervised machine learning-based NLP pipeline to build medical subdomain classifiers that can categorize clinical notes into medical subdomains. Specifically, we compared the performance of various classifiers using different clinical feature representation methods, weighting strategies, and supervised learning algorithms, and we investigated the portability of classifiers across two clinical datasets that we trained classifiers on one dataset and applied directly to the other dataset. We have achieved good accuracy in classifying clinical notes into their medical subdomains.

MATERIALS AND METHODS

Overview

We integrated NLP and other machine learning tools to develop our generalized clinical document classification and prediction pipeline (Figure 1). We used two sets of clinical notes to conduct the study. The datasets were acquired from the Integrating Data for Analysis, Anonymization, and Sharing (iDASH) data repository and Massachusetts General Hospital (MGH) clinical notes in the Research Patient Data Registry (RPDR) data repository of the Partners HealthCare system.[15]

Clinical Dataset

iDASH (Integrating Data for Analysis, Anonymization, and Sharing) Dataset

We downloaded 431 publicly available anonymized clinical notes or reports from the “Clinical Notes and Reports data repository” in the iDASH data repository. The iDASH data repository selected widely diverse clinical notes and reports from MedicalTranscriptionSamples.com, which is a website that collects sample notes and reports from various transcriptionists and clinical users. The iDASH documents include admission notes, discharge notes, progress notes, surgical notes, outpatient clinic notes, emergency notes, echocardiogram, CT scan, MRI, nuclear medicine, radiographs, ultrasound and radiological procedures reports. Two well-trained clinicians independently and manually annotated each document, assigning it to one of six medical subdomains: ‘Cardiology’, ‘Gastroenterology, ‘Nephrology’, ‘Neurology’, ‘Psychiatry’ and “Pulmonary disease”. Cohen’s κ coefficient of 0.97 was obtained, which represented an excellent inter-rater consistency of annotation. These annotations serve as ground truth for our learning methods.

MGH (Massachusetts General Hospital) Dataset

The MGH dataset includes 542,744 clinical notes of 4,844 patients since 2012, who had visited one of three specialist clinics (neurology, cardiology, and endocrinology) at least once in May 2016 at MGH, the tertiary care medical center in Boston, MA. We limited the note extraction query in the three specialties due to the limited data access. To allow derivation of gold standard labels without needing extensive manual annotations, we extracted all specialist-written notes and created an automated mapping script, which allows the mapping between note authors and their medical subdomains using the Partners Enterprise data warehouse (EDW) physician database.

We further removed notes written by specialists with more than one specialty to ensure that each note can be classified into only one medical subdomain. After removing 386,903 notes that did not fulfill the note selection criteria (Supplementary figure 1), we selected the top 24 medical subdomains among 105 medical specialties in the MGH dataset (Supplementary table 1). The remaining 91,237 clinical notes were deidentified by 'deid' software after data filtering,[16,17] and used for the further analysis. The deidentification not only helps to protect the patients' identities but also prevents the classification system from relying on the name of specialists for the classification task because the names are elided. The document filtering process is shown in Supplementary figure 1. The MGH dataset was acquired through Partners Healthcare RPDR system,[15] and was performed under an Institutional Review Board protocol reviewed and approved by Partners HealthCare (P20160011).

Clinical Feature Representation

Appropriate clinical feature representation has been shown to improve the performance of machine learning classifiers.[10] To extract and represent meaningful clinical features, we adopted the clinical NLP annotator and parser, Apache clinical Text Analysis and Knowledge Extraction System (cTAKES),[18] and used the Unified Medical Language System (UMLS) Metathesaurus, and Semantic Network to filter clinically relevant UMLS concepts in clinical notes.[19-21]

We used the bag-of-words representation, which directly identified and normalized lexical variants from the unstructured text content, as the baseline of clinical feature representation. For clinically relevant concept identification, we selected the cTAKES analysis engine, Aggregate Plaintext UMLS Processor, to acquire UMLS concept unique identifiers (CUIs) and build feature sets. The UMLS Metathesaurus and Semantic Network were further applied to restrict the extracted UMLS CUIs within clinically relevant semantic groups and semantic types. We selected 56 semantic types within five clinically related semantic groups, which are "Anatomy (ANAT)", "Chemicals and Drugs (CHEM)", "Disorders (DISO)", "Phenomena" (PHEN) and "Procedures (PROC)". We further restricted UMLS-derived concepts to 15 semantic types (Table 1), which are most related to clinical tasks.

Table 1. Fifteen semantic types selected for clinical feature representation.

TUI	**Semantic group**	**Semantic type description**
T017	Anatomy	Anatomical Structure
T022	Anatomy	Body System
T023	Anatomy	Body Part, Organ, or Organ Component
T033	Disorders	Finding
T034	Phenomena	Laboratory or Test Result
T047	Disorders	Disease or Syndrome
T048	Disorders	Mental or Behavioral Dysfunction
T049	Disorders	Cell or Molecular Dysfunction

T059	Procedures	Laboratory Procedure
T060	Procedures	Diagnostic Procedure
T061	Procedures	Therapeutic or Preventive Procedure
T121	Chemicals & Drugs	Pharmacologic Substance
T122	Chemicals & Drugs	Biomedical or Dental Material
T123	Chemicals & Drugs	Biologically Active Substance
T184	Disorders	Sign or Symptom

We also built three hybrid feature sets using the combination of bag-of-words + UMLS concepts, bag-of-words + UMLS concepts restricted to five semantic groups, comprising 56 semantic types, as well as bag-of-words + UMLS concepts restricted to 15 semantic types. Through NLP, ontology and semantic filtering, clinical knowledge in clinical notes was represented in a uniform way.

For different feature sets, we preserved all of the extracted features instead of applying additional feature selection methods to subset the features. In addition to using the term frequency of features, term frequency–inverse document frequency (tf-idf) weighting is also applied to emphasize the importance of features.[22] All bag-of-words features were processed by word tokenization and word stemming using the Porter stemming algorithm.

Supervised Machine Learning

We constructed 98 binary one-versus-rest classifiers for each set of clinical notes using supervised learning algorithms with five-fold cross-validation and three repetitions. The 98 classifiers include the combinations of seven clinical feature representation methods, two vector representation methods, and seven supervised learning algorithms. We used a multinomial naïve Bayes (NB) algorithm as the baseline algorithm and compared against L1- or L2-regularized multinomial logistic regression, SVM with linear kernel,[23,24] linear SVM with stochastic gradient descent (SGD), and two ensemble algorithms, random forest and adaptive boosting. Classifiers output the class probability of all medical subdomain labels, and the label with the highest probability was regarded as the predicted result and compared against the ground truth label for evaluation.

Evaluation

To evaluate the performance of binary classifiers, we used balanced accuracy ($\frac{1}{2} \times \frac{True\ positive}{All\ positive} \times \frac{True\ negative}{All\ negative}$),[25] precision, recall, F1 score, and area under receiver operating characteristic curve (AUC) as performance metrics. Statistical analyses of unequal variances *t*-tests (Welch's t-test) between groups were used as the significance test.

Tools

The pipeline was built on cTAKES and python version 2.7.11. The Natural Language Toolkit ('nltk') package was used for lexical normalization (word tokenization and stemming process) of bag-of-words features generation, and for the tf-idf weighting adjustment. 'scikit-learn' package was selected for the supervised learning algorithms implementation and model evaluation. Data processing, statistical analysis, and figure generation were done in Python 2.7.11 and R 3.3.2 with customized scripts. The source code of the pipeline is available online at https://github.com/ckbjimmy/cdc/.

RESULTS

Optimized Model for Medical Subdomain Classification

We represented the clinical features in two sets of clinical notes using different feature representation methods (Table 2).

Table 2. Dimension of feature sets using different clinical feature representation.

Dimension of the feature set	iDASH	MGH
Bag-of-words (Vocabulary size)	10150	160097
UMLS concepts	4750	25456
UMLS concepts restricted to five semantic groups	4531	24457
UMLS concepts restricted to 15 semantic types	3634	18520
Bag-of-words + UMLS concepts	14900	185553
Bag-of-words + UMLS concepts restricted to five semantic groups	14681	184554
Bag-of-words + UMLS concepts restricted to 15 semantic types	13784	161949

We combined different clinical feature and vector representation methods with supervised learning algorithms to generate medical subdomain classifiers for clinical notes. The baseline classifier used the bag-of-words, term frequency representation and NB algorithm. In the iDASH dataset, combining the hybrid features of bag-of-words + UMLS concepts restricted to five semantic groups, with tf-idf weighting and linear SVM algorithm yielded the best performing classifier for medical subdomain classification (F1 score of 0.932, AUC of 0.957), followed by using the bag-of-words + all UMLS concepts or using the bag-of-words + UMLS concepts restricted to 15 semantic types as the feature representation with tf-idf weighting and linear SVM algorithm. The classifiers built by these combinations outperformed the baseline classifier with statistical significance ($p < 0.01$) (Table 3, Figure 2 for F1 score, Supplementary figure 2 for AUC).

Table 3. Top five best-performed classifiers in iDASH and MGH datasets.

Data	Feature	Vector	Algorithm	F1	AUC	p-value
iDASH	Bag-of-words + UMLS (5SG)	Tf-idf	SVM-Lin	0.932	0.957	<0.01
	Bag-of-words + UMLS (All)	Tf-idf	SVM-Lin	0.931	0.957	<0.01
	Bag-of-words + UMLS (15ST)	Tf-idf	SVM-Lin	0.930	0.957	<0.01
	Bag-of-words + UMLS (All)	Tf-idf	SVM-Lin-SGD	0.928	0.955	<0.01
	Bag-of-words	Tf-idf	SVM-Lin	0.927	0.955	<0.01
	Bag-of-words	**Tf**	**NB**	**0.893**	**0.935**	**Baseline**
MGH	Bag-of-words + UMLS (5SG)	Tf-idf	SVM-Lin	0.934	0.964	<0.01
	Bag-of-words + UMLS (15ST)	Tf-idf	SVM-Lin	0.931	0.962	<0.01
	Bag-of-words + UMLS (All)	Tf-idf	SVM-Lin	0.930	0.962	<0.01
	Bag-of-words	Tf-idf	SVM-Lin	0.924	0.958	<0.01
	Bag-of-words + UMLS (5SG)	Tf	LR-L1	0.915	0.953	<0.01
	Bag-of-words	**Tf**	**NB**	**0.755**	**0.867**	**Baseline**

Abbreviation: SG: semantic groups, ST: semantic types, Tf: term frequency, Tf-idf: term frequency-inverse document frequency weighting, SVM-Lin: linear support vector machine, SVM-Lin-SGD: linear support vector machine with stochastic gradient descent training, LR-L1: L1-regularized multinomial logistic regression, NB: Multinomial naïve Bayes. Baseline combinations are shown in bold face.

In the MGH dataset, the linear SVM classifier with tf-idf weighting and the hybrid feature representation of bag-of-words + UMLS concepts restricted to five semantic groups also yielded the best performance (F1 score of 0.934, AUC of 0.964), which significantly outperformed the baseline NB classifier with the term frequency and bag-of-words combination (Table 3, Figure 2 for F1 score, Supplementary figure 2 for AUC). Relaxing the semantic feature representation also yielded optimally performing classifiers (figure 2). In general, classifiers constructed by the combination of the hybrid feature representation of bag-of-words + UMLS concepts restricted to five semantic groups or 15 semantic types, with tf-idf weighting representation and linear SVM algorithms yielded better performance on classifying the clinical notes into the correct medical subdomain in both iDASH and MGH datasets.

We further extracted important features by ranking coefficients of variables in the L1-regularized multinomial logistic regression classifier. Top important features of six medical subdomains in the iDASH and MGH classifiers are listed in Supplementary table 2.

Error Analysis

For each dataset, we compared all performance metrics between the baseline and the best-performing classifiers. Balanced accuracies of the baseline and the best classifiers of iDASH dataset are 0.896 and 0.932, respectively, and balanced accuracies of the baseline and the best classifiers of MGH dataset are 0.763 and 0.925, respectively. Regardless of different combinations of the clinical feature representation and machine learning algorithm, the specificity and negative predictive value (NPV) are consistently high. However, the recall (sensitivity) and precision (positive predictive value) are low in some medical subdomains (Figure 3).

The best-performing iDASH and MGH classifiers, which used the hybrid feature representation of bag-of-words + UMLS concepts restricted to five semantic groups, with tf-idf weighting and linear SVM, yielded significant improvement in these three medical subdomains, comparing to the baseline classifiers. Figure 3(a) shows that the precision and F1 score of the baseline iDASH classifier are low in medical subdomains of "Pulmonary disease" (F1 score of

0.749 and precision of 0.667) and 'Nephrology' (F1 score of 0.715 and precision of 0.667). The recall is low in 'Psychiatry' (F1 score of 0.914 and recall of 0.841). In the best iDASH classifier, the F1 score and precision in the medical subdomain "Pulmonary disease" are 0.833 and 0.804, and in 'Nephrology' are 0.857 and 0.818, respectively. The F1 score and recall of 'Psychiatry' are 0.968 and 0.938, respectively. Confusion matrices of classification tasks using the baseline and the best iDASH classifiers are shown in Supplementary table 3.

Figure 3(b) demonstrated that the baseline classifier for the MGH dataset yielded low precision in many medical subdomains. Nine of 24 medical subdomains have precision lower than 0.6 ('Anesthesiology', "General surgery", 'Hematology', "Infectious diseases" "Intensive care", 'Neurosurgery', "Obstetrics and gynecology", 'Otolaryngology' and "Pulmonary disease") and four of 24 medical subdomains have recall lower than 0.6 ("Geriatric medicine", "Medical oncology", 'Pediatrics' and "Pediatric neurology"). The best classifier of MGH data, however, improves most of the measurements to above 0.8, except precision of classifying the "Infectious disease" and "Intensive care" subdomains (precision of 0.797 and 0.776, respectively). F1 score of classifying all medical subdomains are above 0.83.

Model Portability

To examine the model portability across the clinical note datasets, we applied the best classifier of each dataset to classify the medical subdomains in the other dataset. The result shows that the overall accuracy using the best iDASH classifier (with six medical subdomains) to classify medical subdomains of MGH clinical notes is 0.734. The classifier yielded the highest performance in the subdomain 'Cardiology' (F1 score of 0.806, precision of 0.923 and recall of 0.715), and had the lowest performance in the subdomain "Pulmonary disease" with F1 score of 0.307, precision of 0.197 and recall of 0.692. Other subdomains fall in between (Table 4).

Table 4. Model portability test. The performance of using the best iDASH classifier to classify the medical subdomain of MGH clinical notes, and using the best MGH model to classify the medical subdomain of iDASH documents.

From iDASH to MGH					**From MGH to iDASH**				
Subdomain	**AUC**	**Precision**	**Recall**	**F1**	**Subdomain**	**AUC**	**Precision**	**Recall**	**F1**
Cardiology	0.828	0.923	0.715	0.806	**Cardiology**	0.731	0.829	0.500	0.624
Gastroenterology	0.802	0.396	0.691	0.503	**Gastroenterology**	0.832	1.000	0.664	0.798
Neurology	0.877	0.745	0.859	0.798	**Neurology**	0.775	0.902	0.567	0.696
Psychiatry	0.803	0.907	0.613	0.732	**Psychiatry**	0.941	0.794	0.900	0.844
Pulmonary	0.820	0.197	0.692	0.307	**Pulmonary**	0.545	1.000	0.089	0.164
Nephrology	0.770	0.573	0.561	0.567	**Nephrology**	0.634	0.750	0.273	0.400

The overall accuracy of using the best MGH classifier (with 24 medical subdomains) to classify medical subdomains of iDASH notes and reports is 0.520. The medical subdomain 'Psychiatry' had the best classification performance with F1 score of 0.844, precision of 0.794 and recall of 0.900, followed by 'Gastroenterology', 'Neurology', 'Cardiology', 'Nephrology', then "Pulmonary disease".

DISCUSSION

The purpose of the study was to classify the medical subdomain of an unstructured clinical note accurately, and we demonstrated that the machine learning-based NLP approach could be a solution for building portable medical subdomain classifiers for clinical notes. Using two sets of clinical notes, we found that the selection of a classifier-building combination of the clinical feature representation and supervised learning algorithm is important to yield a better-performing and portable medical subdomain classifier for clinical notes.

Among 98 classifiers with different classifier-building combinations of clinical feature representation and learning algorithms, the classifier constructed by the combination of bag-of-words + UMLS concepts restricted to semantic groups or semantic types as the clinical feature, with tf-idf weighting and linear SVM algorithm outperformed other combinations in both the iDASH and MGH clinical note datasets. For clinical feature representation, Yetisgen-Yildiz et al. also achieved the best model performance using the word and phrase hybrid approach for clinical note classification.[10] We also adopted the similar bag-of-words and UMLS concept hybrid, which allows us to capture important tokenized words and medical phrases that can't be identified in concepts-only or words-only models. For example, combined features identify both the word 'heart' and the concept "congestive heart failure" when "congestive heart failure" appears in the text. The word 'heart' and the phrase concept "congestive heart failure" are both important features for a cardiology note, yet concepts-only models would identify "congestive heart failure" while words-only models would identify 'heart' and miss the full concept "congestive heart failure".

Adding UMLS concepts restricted to semantic groups or semantic types on the basis of the bag-of-words feature slightly augments the classifier performance, yet using the bag-of-words feature is necessary to yield the optimal result. Semantic restriction reduces the size of the feature space by removing clinically irrelevant concepts and therefore decreases the model complexity. However, the bag-of-words feature includes some words, which may not be recognized as medical concepts by clinical NLP systems (e.g. abbreviation, neologism), but would be important for identifying the medical subdomain of a clinical document. Therefore, combining the bag-of-words feature with semantic restricted medical concepts is useful to compensate for the disadvantages of missing those words in the pure concept approach.

In our study, SVM with linear kernel outperformed other supervised learning algorithms, and was followed by regularized multinomial logistic regression. The result shows that the algorithm selection is consistent with previous studies, in which SVM with linear kernel is known as an effective model for high dimensional datasets, and D'Avolio et al. adopted multinomial logistic regression (maximum entropy algorithm) in the ARC system to achieve good performance for the image and pathology report classification task.[14] To minimize the effect of model overfitting and model instability, repeated five-fold cross-validation was adopted in all modeling processes. Binary one-versus-rest classifiers rather than multi-class classifiers were used to reduce the evaluation complexity.

Many specific medical subdomains, such as 'Psychiatry' and 'Neurology', yielded good performance and portability across clinical datasets. However, some paired medical subdomains such as "Pulmonary disease" and 'Nephrology' are difficult to distinguish by classifiers because they usually share patients with similar clinical conditions. In the iDASH classifiers, we found that the subdomains "Pulmonary disease" and 'Nephrology' have lower precision, and 'Cardiology' has relatively poor recall. This may imply that some pulmonology and nephrology cases are misclassified to cardiology. The possible cause is that patients in pulmonology and nephrology clinics may share the same features, such as dyspnea, with patients in cardiology clinics. Overlapping features lead to a harder classification task between these medical subdomains. The relatively poor performance in 'Anesthesiology', "Infectious disease", and "Intensive care" subdomains can also be explained by the patient similarity with other subdomains. By contrast, certain medical subdomains, for example, 'Neurology', "Orthopedic surgery", 'Psychiatry', "Radiation oncology", and 'Urology', usually yield better performance because of the uniqueness of their features.

Important features of classifiers are useful for clinicians to understand how the classifier makes the decision. It can also be used for developing a domain ontology for NLP-driven research in specific medical domains.[26] We identified the top features of different medical subdomains, but some ambiguous or clinically unrelated words and phrases also appear on the list, which indicates that the classifier fitted not only meaningful data but also noise. We also found that the important features in different datasets are both meaningful but varied. Table 4 shows that the number of overlapped features is limited. This is because the characteristics of two sets of clinical notes are different. Notes and reports in the iDASH dataset include outpatient notes, inpatient summaries, procedure reports, and examination reports, yet MGH clinical notes are mainly outpatient notes. The suboptimal performance of the MGH classifier portability also revealed the issue that the content of the MGH dataset is more homogeneous in comparison with the iDASH dataset. To achieve better performance of model portability and to build generalizable classifiers, source and target data may need to have similar features.

The strength of the study is that we took advantage of the combination of hybrid clinical knowledge representation methods and supervised machine learning algorithms for medical subdomain classification of clinical notes, which has not been explored extensively. We also used standardized terminology in the UMLS Metathesaurus for clinical feature representation, and we further identified clinically relevant UMLS concepts using semantic groups and semantic types in the Semantic Network. Using standardized terminology can be a good knowledge representation approach, which also provides the possibility of future clinical EHR system integration.

There are also some limitations of the study. First, we only adopted the NLP analysis tools from cTAKES. We did not examine other clinical NLP systems for performance comparison. Though cTAKES includes an NLP pipeline with promising performance,[18] there are still other options, such as MetaMap from National Library of Medicine (NLM),[27] the Clinical Language Annotation, Modeling and Processing Toolkit (CLAMP) developed by the NLP team at The University of Texas Health Science Center at Houston, and the name entity-specific tool Clinical Named Entity Recognition system (CliNER).[28] Further investigation on different clinical NLP systems is required to understand whether cTAKES is the most suitable tool for use in predicting the medical subdomain of a clinical document. Additionally, we investigated only two clinical note datasets. To be generalizable, further investigation on more datasets is required. We also found that a few physicians' first names appear in our feature spaces of MGH classifiers, which indicates that the process of deidentification was not perfect. Further improvement of deidentification is still required to prevent classification tasks from using the information of specific healthcare providers. For example, using deep learning to replace the current dictionary-based approach might improve performance.[29] Finally, we would need to do additional external validation by experienced clinicians to integrate the medical subdomain classification into real-world clinical decision support system.

The machine learning-based NLP approach to classify the medical subdomain of a clinical note may assist clinicians to redirect patient's unsolved problems to adequate medical specialties and experts in time purely based on the content of clinical notes. Often clinicians encounter patients' clinical problems and dilemmas beyond their domain of expertise, which may leave questions unanswered, and result in misdiagnosis, delayed clinical care, delayed or failure to refer and even lead to inappropriate treatment and management.[30] Identifying the medical subdomain of a clinical note can also help with NLP. For example, the subdomains may generate topics, and topics may generate concepts, phrases and words via generative models for further NLP applications. We plan to integrate the information of both medical subdomain and clinical expert to build hierarchical models to improve our methods, and may adopt domain adaptation and transfer learning techniques to improve the performance of model portability and construct a generalizable solution.

Acknowledgement

This work was supported in part by grants from National Library of Medicine, National Institutes of Health, Grant 5R00LM011575. We are grateful to Dr. William Lester for his suggestions and comments on defining the problem. The authors thank Enterprise Research Infrastructure & Services at Partners Healthcare for their support and computing service. The iDASH dataset used in this study was downloaded from the iDASH repository supported by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54HL108460.

Author Contribution

WWH acquired the data, designed the experiment, implemented the programming tasks, performed the analysis and drafted the manuscript. KBW helped on study design, provided feedback on the data analysis and revision of the manuscript. ATM provided the expertise in NLP and ontology, and critical revision of the manuscript. PSZ supported the design and analysis of machine learning tasks, provided the servers for experiments, and revised the manuscript. HCC supervised the project, helped acquire the data, defined the clinical problems and applications, interpreted data and revised the manuscript. All authors contributed to discussions regarding the interpretation of the results, and agreed with the content of the manuscript.

Reference

1. Sebastiani F. Machine Learning in Automated Text Categorization. ACM Computing Surveys (CSUR) 2002;31(1):1-47.
2. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014. https://www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf. Accessed February 18, 2017.
3. Bernhardt PJ, Humphrey SM, Rindflesch TC. Determining prominent subdomains in medicine. AMIA Annu Symp Proc 2005;46-50.
4. Torii M, Wagholikar KB, Liu H. Using machine learning for concept extraction on clinical documents from multiple data sources. J Am Med Inform Assoc 2011;18(5):580-587.
5. Coden A, Savova G, Sominsky I, et al. Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model. J Biomed Inform 2008;42:937-949.

6. Wagholikar KB, MacLaughlin KL, Henry MR, et al. Clinical decision support with automated text processing for cervical cancer screening. J Am Med Inform Assoc 2012;19(5):833-839.

7. Byrd RJ, Steinhubl SR, Sun J, et al. Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records. Int J Med Inform 2014;83(12):983-992.

8. Liao KP, Ananthakrishnan AN, Kumar V, et al. Methods to Develop an Electronic Medical Record Phenotype Algorithm to Compare the Risk of Coronary Artery Disease across 3 Chronic Disease Cohorts. PLoS One 2015;10(8):e0136651.

9. McCoy TH, Castro VM, Cagan A, et al. Sentiment Measured in Hospital Discharge Notes Is Associated with Readmission and Mortality Risk: An Electronic Health Record Study. PLoS One 2015;10(8):e0136341.

10. Yetisgen-Yildiz M, Pratt W. The effect of feature representation on MEDLINE document classification. AMIA Annu Symp Proc 2005;849-53.

11. Tsatsaronis G, Macari N, Torge S, et al. A Maximum-Entropy approach for accurate document annotation in the biomedical domain. J Biomed Semantics 2012;(3):Suppl 1:S2.

12. Friedman C, Alderson PO, Austin JH, et al. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1994;1(2):161-174.

13. Wilcox A, Hripcsak G, Friedman C. Using Domain Knowledge Sources to Improve Classification of Text Medical Reports [poster]. KDD-2000 Workshop on Text Mining 2000.

14. D'Avolio LW, Nguyen TM, Farwell WR, et al. Evaluation of a generalizable approach to clinical information retrieval using the automated retrieval console (ARC). J Am Med Inform Assoc 2010;17:375-382.

15. Murphy SN, Chueh HC. A security architecture for query tools used to access large biomedical databases. Proc AMIA Symp 2002;552-556.

16. Neamatullah I, Douglass MM, Lehman LW, et al. Automated de-identification of free-text medical records. BMC Med Inform Decis Mak 2008;(8):32.

17. Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215-e220.

18. Savova GK, Masanz JJ, Ogren P V, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Informatics Assoc 2010;17(5):507-513.

19. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(90001):D267-270.

20. McCray AT. An upper-level ontology for the biomedical domain. Comp Funct Genomics 2003;4(1):80-84.

21. McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Stud Health Technol Inform. 2001;84(Pt 1):216-20.

22. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Information Processing & Management 1988;24(5):513-523.

23. Cortes C, Vapnik V. Support-Vector Networks. Machine Learning 1995;20(3):273-297.

24. Fan RE, Chang KW, Wang XR, et al. LIBLINEAR: A Library for Large Linear Classification. Journal of Machine Learning Research 2008;9:1871-1874.

25. Brodersen KH, Ong CS, Stephan KE, et al. The balanced accuracy and its posterior distribution. Proceedings of the 20th International Conference on Pattern Recognition. IEEE Computer Society 2010;3121-3124.

26. Musen MA. Domain ontologies in software engineering: Use of Protégé with the EON architecture. Methods of Information in Medicine 1998;37(4-5):540-550.

27. Aronson AR. Effective Mapping of Biomedical Text to the UMLS Metathesaurus: The MetaMap Program. Proc AMIA Symp 2001;17-21.

28. Boag W, Wacome K, Naumann T, et al. CliNER: A Lightweight Tool for Clinical Named Entity Recognition [abstract]. AMIA Joint Summits on Clinical Research Informatics 2015.

29. Dernoncourt F, Lee JY, Uzuner O, et al. De-identification of patient notes with recurrent neural networks. J Am Med Inform Assoc 2016;ocw156. doi.org/10.1093/jamia/ocw156

30. Weingart SN, Ship AN, Aronson MD. Confidential clinician-reported surveillance of adverse events among medical inpatients. J Gen Intern Med 2000;15(7):470-477.

Legend

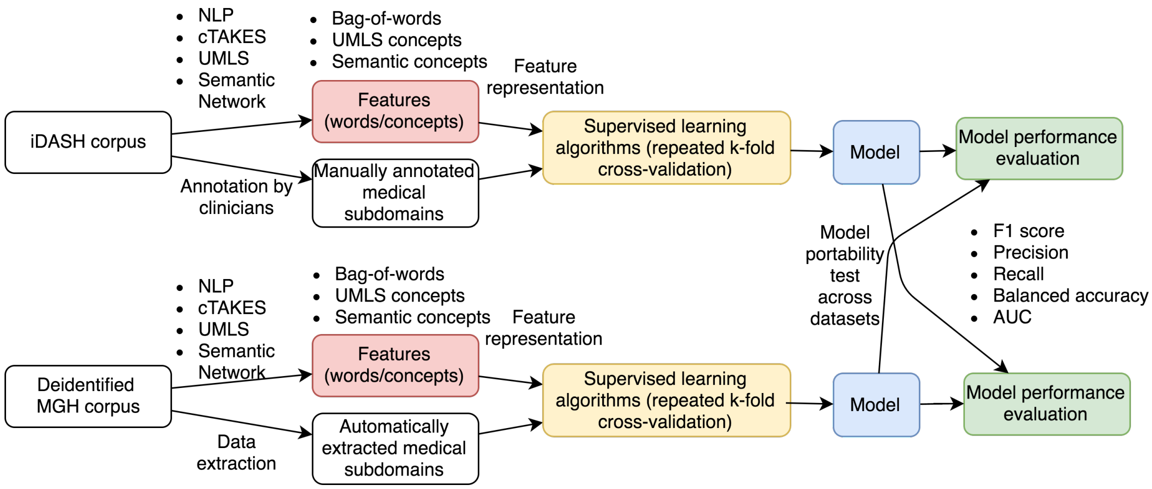
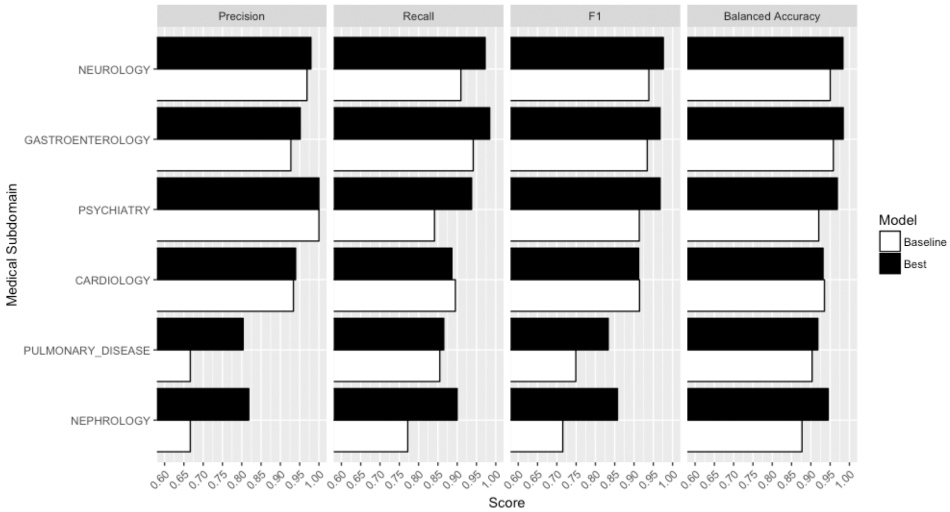


Figure 1. The study design. We used two datasets — clinical notes and reports from the Integrating Data for Analysis, Anonymization, and Sharing (iDASH) data repository as well as Massachusetts General Hospital (MGH) clinical notes from the Research Patient Data Registry (RPDR) data repository of the Partners HealthCare system. For each dataset, we applied and combined different clinical feature representation methods, weighting strategies, and supervised learning algorithms to build classifiers. F1 score, precision, recall, balanced accuracy and area under receiver operating characteristic curve (AUC) were used to evaluate the model performance. The model portability test across datasets was performed. We have applied the clinical NLP system, clinical Text Analysis and Knowledge Extraction System (cTAKES), the UMLS Metathesaurus, Semantic Network, and machine learning tools to construct the pipeline. The analytic pipeline has three main components, the medical concept extractor (red), model constructor (yellow), and evaluator (green).

Figure 2. The performance of classifiers (using F1 scores) built by different combinations of the clinical feature and vector representation method with supervised learning algorithm. In both sets of clinical notes, the combination of the hybrid features of bag-of-words + UMLS concepts restricted to five semantic groups with tf-idf weighting and linear SVM yielded the optimal performance for clinical note classification based on the medical subdomain of the document. (a)

F1 score of classifiers trained on iDASH dataset, (b) F1 score of classifiers trained on MGH dataset. The lines connecting data points for different clinical feature representation methods only serve to tie together the visual results from specific algorithms on different sets of features, but should not imply continuity in the horizontal axis features.

(a) Performance Comparison in iDASH Dataset
- Bag-of-words + Frequency Count + Naive Bayes (Baseline)
- Bag-of-words with UMLS concepts (5 Semantic Groups) + tf-idf weighting + linear SVM (Best)



(b) Performance Comparison in MGH Dataset
- Bag-of-words + Frequency Count + Naive Bayes (Baseline)
- Bag-of-words with UMLS concepts (5 Semantic Groups) + tf-idf weighting + linear SVM (Best)

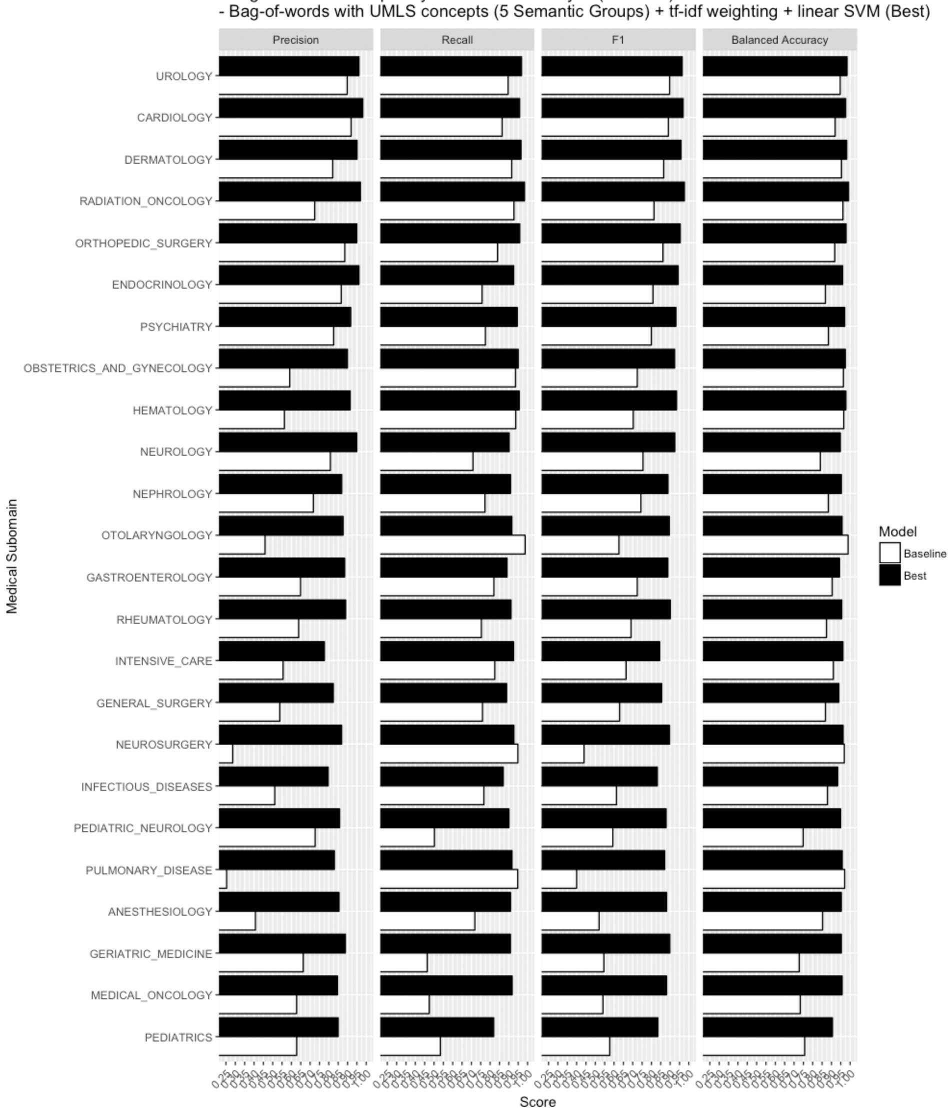
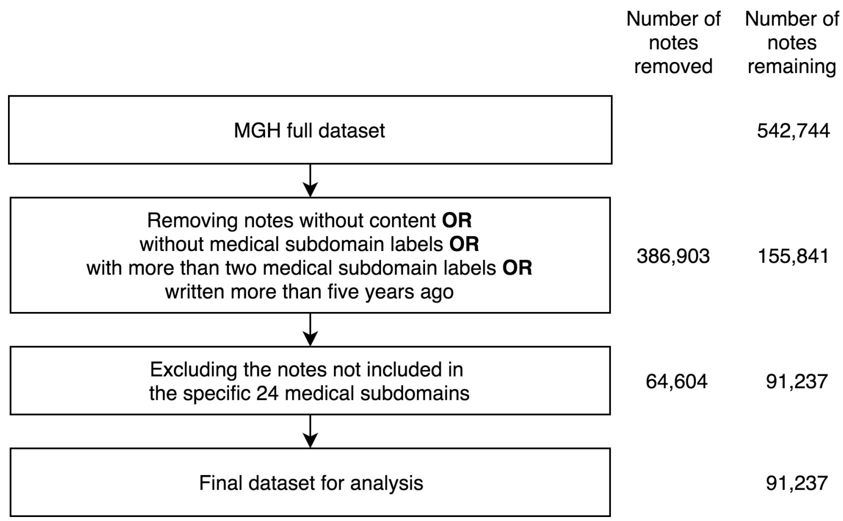


Figure 3. The performance across different medical subdomains in the baseline and the best classifiers on iDASH and MGH datasets. All measurements, including precision, recall, F1 score, balanced accuracy, and AUC were compared in the (a) baseline (white) and the best (black) iDASH classifiers, and the (b) baseline (white) and the best (black) MGH classifiers. Significantly improved performance is observed in the best classifier, especially in difficult to separate medical subdomains, such as ‘Anesthesiology’, “Pulmonary disease”, “Intensive care” and “Infectious diseases”.

Appendix



Supplementary figure 1. The Final Dataset Selection Process of MGH Dataset.

Supplementary figure 2. The performance of classifiers (using AUC) built by different combinations of the clinical feature representation method, vector representation method and supervised learning algorithm. In both datasets, the combination of the hybrid feature of bag-of-words + UMLS concepts restricted to five semantic groups with tf-idf weighting and linear SVM yielded the optimal performance for clinical note classification based on the medical subdomain of the document. (a) AUC of classifiers trained on iDASH dataset, (b) AUC of classifiers trained on MGH dataset. The lines connecting data points for different clinical feature representation methods only serve to tie together the visual results from specific algorithms on different sets of features, but should not imply continuity in the horizontal axis features.

Supplementary table 1. Representative medical subdomains in the iDASH and MGH dataset. We selected the top 24 medical subdomains from 105 medical specialties in the MGH dataset.

Medical Subdomain	**Number of Documents (iDASH)**	**Number of Documents (MGH)**
Cardiology	116	20,928
Endocrinology	-	12,395
Neurology	97	10,974
Pediatrics	-	4,790
General surgery	-	4,388
Dermatology	-	4,067
Psychiatry	30	3,734
Gastroenterology	110	3,188
Orthopedic surgery	-	3,053
Geriatric medicine	-	2,092
Urology	-	2,090
Anesthesiology	-	1,979
Nephrology	22	1,936
Medical oncology	-	1,881
Obstetrics and gynecology	-	1,784
Infectious diseases	-	1,729
Pediatric Neurology	-	1,655
Rheumatology	-	1,536
Otolaryngology	-	1,473

Radiation oncology	-	1,445
Neurosurgery	-	1,414
Hematology	-	1,036
Intensive care	-	907
Pulmonary disease	56	763
Total	431	91,237

Supplementary table 2. Ranked top important features (post-stemming, bag-of-words + UMLS concepts restricted to five semantic groups) of six medical subdomains identified by iDASH and MGH classifiers. The phrases in the parentheses are the UMLS descriptions of the corresponding UMLS CUIs.

<table>
<tr><th>Top Features in iDASH Model</th><th>Top Features in MGH Model</th></tr>
<tr><td colspan="2">CARDIOLOGY</td></tr>
<tr><td>blood | perform | bypass | systol | eject | diagnosis:1. | vein | arrest.2 | c0558145 (Skin appearance normal (finding)) | diabet | c0817096 (Chest) | insert | done | disease.3 | beat | mitral | pain | c0020538 (Hypertensive disease) | pacemak | doxycyclin | c0232201 (Sinus rhythm) | left | c0013516 (Echocardiography) | obes | c1269008 (Entire coronary artery) | c0205042 (Coronary artery) | c0003842 (Arteries) | chest | doe | palpit | valv | sinu | studi | rhythm | minut | follow | arteri | aortic | rate | wire | lead | dr. | subclavian | c1281570 (Entire heart) | c0018787 (Heart) | atrial | ventricular | coronari | heart | cardiac</td><td>reinforc | c0428474 (Serum LDL cholesterol measurement) | c0428897 (Jugular venous pressure) | reaction | mba | casresultsreportsnot | select | c0226896 (Oral | cavity) | facc | transcrib | prophylaxi | somat | kind | cce | confirm | shx | arbor | c0085619 (Orthopnea) | c0200045 (Manual pelvic examination (procedure)) | nsca | pelagia | c1623258 (Electrocardiography) | oht | recreat | bi | c0278005 (Normal bowel sounds) | beeper | ido | present | mese | statu | pmi | c0400018 (Diagnostic endoscopic examination on colon) | parasthesia | habitsrisk | document | disposit | interv | educationcounsel | pshx | misaglign | planter | narr | c1287400 (History finding) | c0457086 (Morning stiffness - joint) | jvp | electron | fisher | c0013146 (Drug abuse) | cholesterolldl</td></tr>
</table>

GASTROENTEROLOGY	
c0014876 (Esophagus) | c1278919 (Entire esophagus) | murmur | vomit | mucosa | c0009378 (colonoscopy) | gallbladd | pancrea | distal | moder | sever | transfer | c1278925 (Entire cecum) | c0038351 (Stomach) | c0007531 (Cecum) | c0009368 (Colon structure (body structure)) | portion | duodenum | rectum | duct | rectal | stool | given | appendix | c0021853 (Intestines) | posit | advanc | endoscopi | diet | colonoscop | c0000726 (Abdomen) | liver | dilat | stomach | nausea | pelvi | abdomen | lesion | cholesterol | also | discuss | procedur | colonoscopi | cecum | bowel | esophagu | without | polyp | abdomin | colon	ibd | le | precancer | coliti | c0009378 (colonoscopy) | abduct | ppi | relax | rheum | methocarbamol | c1457887 (Symptoms) | c0231377 (At risk for impaired home maintenance management) | hypercholesterolemia | hcc | sptrg | thiim | esophagu | c0021853 (Intestines) | formalin | c0392916 (Intracellular ferritin) | cmd | manometri | constip | mrn | perin | c0023895 (Liver diseases) | stool | c0018834 (Heartburn) | motil | c0014245 (Endoscopy (procedure)) | c0719635 (DOS brand of docusate sodium) | endoscop | c0201539 (Alpha one fetoprotein measurement) | djd | colon | c1299487 (Patient name) | crohn | motion | liver | egd | c0193388 (Biopsy of liver (procedure)) | outsid | tel | impressionplan | c0221565 (Encounter due to family history of arthritis) | perian | gastroenterolog | hsm | allostat
NEPHROLOGY	
red | longitudin | go | c0227613 (Right kidney) | recent | check | echotextur | bout | secur | hemodialysi | problem | ani | tie | hypertens | protein | transplant | c0227614 (Left kidney) |	uaurobi | agre | c0242429 (Sore Throat) | protein | cr | nsaid | c0031140 (Peritoneal Dialysis, Continuous Ambulatory) | egfr | kidney | stiff | c0426663 (Abdomen soft) |

c0020295 (Hydronephrosis) | c0555903 (Total protein measurement) | histori | glucos | hi | hematuria | bladder | c0022661 (Kidney Failure, Chronic) | size | ultrasound | c0022658 (Kidney Diseases) | postvoid | failur | dissect | c1278978 (Entire kidney) | hydronephrosi | measur | promis | c0203408 (Echography of kidney) | ureter | clear | daili | approxim | cyst | discharg | c0022646 (Kidney) | blood | cell | diseas | creatinin | urin | kidney | renal	proteinuria | c0019360 (Herpes zoster disease) | kalim | prograf | spgr | pager | cellcept | dip | amlodipin | lcsw | urin | c1533720 (Prednisone 5 MG) | incl | c0205180 (Anicteric) | c0019004 (Hemodialysis) | esrd | sediment | dialysi | disc | temperatur | c0040739 (Transplantation, Homologous) | sed | una | renal | thyromegali | nephrolog | cor | bipolar | ckd | split | msw | physiolog | adenopathi | simic | ext |
NEUROLOGY	
matter | c0024485 (Magnetic Resonance Imaging) | tone | region | sensori | hand | c0016928 (Gait) | sleep | c0013819 (Electroencephalography) | husband | dure | hi | gait | c0228174 (Cerebral hemisphere structure (body structure)) | tumor | episod | tempor | awak | movement | craniotomi | speech | memori | consist | clinic | gener | bilater | intact | unremark | hematoma | cerebr | throughout | mri | nerv | huntington | note | muscl | motor | weak | symmetr | eeg | head | frontal | neurolog | thi | brain | subdur | record | seizur | headach | activ	tcd | donepezil | comprehens | dilut | nystagmu | c0700594 (Radiculopathy) | c0013362 (Dysarthria) | mrcp | yearold | coher | leg | movement | brain | icu | c0027853 (Neurologic Examination) | wl | drive | lifethreaten | c0013839 (Electromyography) | cognit | swing | c0064636 (lamotrigine) | softwar | drift | c0460002 (body system) | neurooncolog | c0026650 (Movement Disorders) | botulinum | righthand | saccad | exmnd | epilepsi | wac | flexor | stroke | ivig | cheng | neuromuscular | emotionallytrigg | zelim | phd | neuropsychiatri | neurolog | amnest | c0150173 (Cognitive

	restructuring) | msph | funduscop | neurocrit |
PSYCHIATRY	
time | orient | c0438696 (Suicidal) | development | hospit | contact | c0033975 (Psychotic Disorders) | c0018524 (Hallucinations) | need | feel | affect | c0344315 (Depressed mood) | famili | data | thought | seroquel | laboratori | patient | bipolar | deferred.axi | live | father | discharg | appropri | iii | hallucin | c0004457 (Axis vertebra) | diagnos | mother | substanc | seclus | p.o | unknown | abus | problem | axi | year | psychosi | unabl | secondari | mental | depress | deni | treatment | disord | medic | psychiatr | mood | quot | behavior	psychopharmacolog | citalopram | c0004457 (Axis vertebra) | retrain | director | unabl | lorazepam | licsw | registr | haldol | suicid | lexapro | report | memori | wish | c0033573 (Psychotic Disorders) | nasosept | card | psychiatr | c0442967 (Salvage procedure) | psych | xanax | discontinu | c0344211 (Supportive care) | sertralin | qh | c0267244 (Right-sided displacement of abomasum) | c0565867 (delivery method) | genitourinari | code | mental | thought | mood | span | factor | ect | gleason | session | abirateron | secur | insight | c0008487 (Chordoma) | judgment | adt | psychiatry31695 | psychiatri | waterfront | axi | mse
PULMONARY | DISEASE	
c0458827 (Airway structure) | flexibl | c0010200 (Coughing) | shunt | nurs | c0032285 (Pneumonia) | stent | scan | room | main | babi | c1962945 (Radiographic imaging procedure) | c1306645 (Plain x-ray) | puls | cpr | found | volum | need | day | bronchoscop | x-ray | requir | trachea | system | wheez | satur | secret |	short | instil | yearold | p16 | c0700198 (Pulmonary aspiration) | region | gtube | c0032227 (Pleural effusion disorder) | mrgc | bipap | lpm | approxim | t2n2b | cough | osa | fev1 | s1s2 | fellow | c0017168 (Gastroesophageal reflux disease) | medicin | scc | advair | nondistend | advis | director |

<table>
<tr>
<td>respiratori | lavag | airway | daughter | short | cough | c0006290 (Bronchoscopy) | diseas | tube | c1278908 (Entire lung) | c0024109 (Lung) | pneumonia | capac | lobe | upper | oxygen | chest | improv | breath | predict | bronchoscopi | lung | pulmonari</td>
<td>wheez | attest | satur | lung | c0002736 (Amyotrophic Lateral Sclerosis) | c0040715 (Chromosomal translocation) | c0226958 (Root of tongue) | lymphadenopathi | dob | history | hospit | sputum | c0022688 (Natural Killer Cells) | dk4875 | nitrolingu | c0035239 (Respiratory Therapy) | ahi | air | fev1fvc | bl3106 | c0235592 (Cervical lymphadenopathy) | dk2130 | c0590708 (Nitrolingual) | c0221910 (Squamous Epithelial Cells) | pulmonari | jugular</td>
</tr>
</table>

Supplementary table 3. The confusion matrices of the classification tasks using the (a) baseline and (b) the best iDASH classifiers.

(a)

Truth \ Predicted	Cardiology	Gastroenterology	Neurology	Psychiatry	Pulmonary	Nephrology
Cardiology	325	0	4	0	12	7
Gastroenterology	8	306	7	3	3	3
Neurology	0	0	282	6	3	0
Psychiatry	0	0	0	90	0	0
Pulmonary	27	6	15	5	112	3
Nephrology	3	13	2	3	1	44

(b)

Truth \ Predicted	Cardiology	Gastroenterology	Neurology	Psychiatry	Pulmonary	Nephrology
Cardiology	327	1	2	0	15	3
Gastroenterology	10	314	1	0	5	0
Neurology	0	0	285	6	0	0
Psychiatry	0	0	0	90	0	0
Pulmonary	25	0	5	0	135	3
Nephrology	7	4	0	0	1	54