Mortality Prediction on ICU Data from MIMIC-III Database

Rojin Aliehyaei, MS
Georgia Institute of Technology, Atlanta, GA

Abstract

This project focuses on delivering a reliable mortality prediction model for patients in the intensive care unit (ICU) using a combination of structured and latent topic features. The meaningful topics relevant to each clinical note are extracted from free-text hospital notes using Latent Dirichlet Allocation model. The results produced a variety of features, where some of which were not utilized in the seminal work by Ghassemi et al. In addition, several binary classifiers, such as Logistic Regression (LR), Linear Support Vector Machines (SVM), Random Forest (RF), and Gradient Boosted Trees (GBT), are implemented using Scala for prediction task, enabling comparative performance evaluation that was not a part of the seminal work, which only utilized the Linear SVM classifier. The final models are evaluated through the area under the receiver operating characteristic curve (AUROC) and area under precision-recall curve (AUPR) using both training and testing data derived from the ICU data in the MIMIC-III database. The results show there are meaningful knowledge that can be extracted from free-text clinical notes which help improving the performance of the mortality prediction model significantly. In addition, GBT with a combined structured and latent topic features creates the most accurate mortality prediction model over LR, Linear SVM, and RF classifiers.

Link to presentation: https://tinyurl.com/yat4d4kj

Introduction

Improving mortality prediction for patients in the intensive care unit (ICU) is an important research topic because 20% of more than 4-million Americans annually admitted into ICU in the United States died while receiving treatments[1]. Additionally, 22% of the hospital cost came from ICU while only 10% of hospital's beds are in the ICU[2]. Thus, accurate mortality prediction models can help saving lives by determining the severity of sickness and evaluating the influences of new healthcare policies as well as treatments in the ICU[3].

Modern technology in Data Science has created new opportunities to improve healthcare industry in recent years. Adoption of electronic health record systems provides an instant access to a large volume of patient records that have been collected over a long period of time. Access to such data by authorized users helps identifying circumstances that influence the condition of patients. Several prior research works focused on building mortality prediction models. However, most of these studies are based on the numeric and snapshot data. In addition to structured data stored in standardized format, there are free-text notes that contain clinical analytics and additional information about each patient. The free-text notes recorded by healthcare experts can provide valuable knowledge for building mortality prediction models. This project focuses on using both structured and unstructured clinical notes from the MIMIC-III database to build a mortality prediction model.

Related Work

A survey on mortality prediction models and tools shows there are a variety of research works in this area[4]. It concludes that most of the models and/or tools have modest accuracy with large variations in their predictive accuracies. Clinical notes, such as discharge summaries or data gathered from 24 hours of intensive care unit (ICU) stay, helped to build a model that predicts mortality with a higher acuity score[5,6]. Several researchers focused on building a model that is enriched with information obtained from free-text notes. Specifically, Frideman et al. used Natural Language Processing (NLP) for extracting information from free-text notes to encode clinical documents[7]. A similar work by Solt et al. also used NLP techniques on clinical notes, achieving a higher precision with performance subjecting to the complexity level of language structures in such notes[8]. Saria et al. combined structured data with information obtained from discharge notes to build a mortality prediction model, improving shortcoming of prediction models based entirely on NLP[9]. Similarly, Arnold and Speier focused on identifying predefined context clinical events in a cohort of patients with brain cancer, learning the temporal patterns between topics in patient records that later applied to predict mortality[10]. Finally, Lehman et al. improved the average AUROC for mortality prediction by using notes written by nurses during the first 24 hours of the ICU stay[11].

This project is inspired by research work on the MIMIC-II (Medical Information Mart for Intensive Care) dataset by Ghassemi et al[5]. Specifically, this project utilizes MIMIC-III dataset[12] to study and evaluate the use of latent variable models for extracting topics from the clinical notes to improve mortality prediction model. The differences between this project and the seminal work by Ghassemi et al.[5] are the utilization of new features, such as admission type, various severity scores, and two binary features related to patient's stay at the hospital and in the ICU. Patients from all age ranges are considered in this study and no patients are excluded. Moreover, other classifiers in addition to Linear SVM (the only classifier used in Ghassemi et al.[5]) are deployed for the prediction task.

Problem Formulation

The main objective of this project is developing a predictive model/classifier that categorizes patients into two groups: patients who died in the hospital and those who survived (i.e., alive).

To construct a predictive model for mortality prediction, two groups of features are utilized: (a) The structured features including age, gender, type of admission, severity scores, and two binary features signifying whether a patient stayed at the hospital or in the ICU for the first time. This group of features are used in the baseline model; (b) Latent topic features obtained from the free-text hospital notes using clustering algorithm. This group of features are used for improving the baseline model.

Area under the receiver operating characteristic curve (AUROC) measures comparative performance of the predictive models under various sets of features.

Data

MIMIC-III is a freely accessible database containing detailed data points about patients admitted to the ICU at the Beth Israel Deaconess Medical Center (BIDMC)[12]. MIMIC database has been widely adopted for clinical research and education around the world. Admission to ICU file contains 46,520 records of distinct patients. The median and average length of stay in the ICU is about 2.1 days and 4.9 days, respectively. Approximately, 44% of patients are female and 56% are male. The ratio of alive to dead patients is approximately 66% to 34% . There are 2,083,180 records in clinical notes of 46,146 distinct patients and 58,361 distinct admissions to the hospital. Almost 900 of these records are flagged as erroneous records. Approximately 12% of these valid note records are related to patients that were not admitted to the ICU. There are 15 categories of clinical notes. The ratio of these categories is shown in Figure 1. Clinical notes labeled as *Discharge summary*, which are accounted for about 3% of all records, are excluded in this study. The reason is that discharge summaries usually contain information about the patient's outcome. The majority of clinical notes are categorized as *Nursing/other* representing about 39% of all valid data.

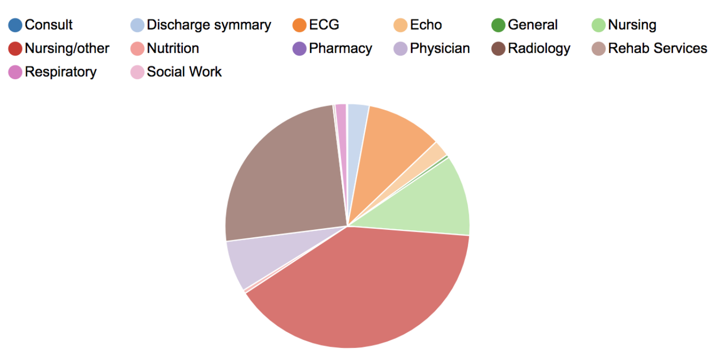

Figure 1: Various categories of clinical notes

Approach and Implementation

As discussed in the prior subsection, two groups of features including structured and latent topic features are used for constructing the predictive model.

The structured features are age, gender, admission type, first hospital stay, first ICU stay, and severity scores including Simplified Acute Physiology Score II (SAPS II), Sequential Organ Failure Assessment (SOFA), Acute Physiology Score III (APS III), and Oxford Acute Severity of Illness Score (OASIS). The categorical features, such as gender and admission type, are transformed to numeric features. Admission type is assigned the number 0, 1, or 2, which indicates whether a record is related to an unplanned medical care (emergency or urgent), the patient's birth (newborn) or a previously planned (elective) admission, respectively. First hospital stay and first ICU stay for the current hospitalization are binary features signifying whether a particular patient's record belongs to the first hospital or the first ICU stay, respectively.

The severity scores upon admission are calculated using various data provided in MIMIC-III database. For calculating SAPS II, the information about patients including ventilation, urine, duration of mechanical ventilation, vital signs, Glasgow Coma Score (GCS), lab results, and blood gases and chemistry values are required. Such data are obtained by querying from various tables including *patients*, *icustays*, *admissions*, *outputevents*, *chartevents*, *diagnoses_icd*, *services*, *labevents*, and *procedureevents*. To calculate SOFA, additional tables, such as *noteevents*, *inputevents_cv*, *inputevents_mv* and *echo* are created. APS III and OASIS values are calculated using the tables created earlier. All severity scores are obtained from the first day of patient's stay in the ICU. Before loading these data in Spark SQL, extra quotation marks are removed from the lab results data. Additionally, in each query the erroneous records are excluded.

A few methods for matching the text against a regular expression were defined. Specifically, for building the echocardiography table, called *echo*, note's timestamp and weight of each patient are obtained by applying user defined functions on the corresponding clinical notes. The LDA optimizer is set to default option, which uses Expectation-Maximization (EM), and its implementation is based on paper by Asuncion et al[13].

Latent topic features are extracted from clinical notes using Latent Dirichlet Allocation (LDA). The input for LDA model is obtained from Term Frequency-Inverse Document Frequency (TF-IDF) technique[14]. It identifies the most informative words for each patient's notes. The TF-IDF measure is the product of TF and IDF as shown below.

$DF(t, D)$: Number of documents in D that contains term t

$|D|$: Total number of documents in the corpus D

$$TF{-}IDF(t, d, D) = TF(t, d).IDF(t, D) \quad (1)$$

$$IDF(t, D) = \log \frac{|D|+1}{DF(t,D)+1} \quad (2)$$

Term frequency or TF(t, d) shows the count of word t in document d. Inverse document frequency or IDF(t, D) is a numerical value that signifies the magnitude of importance for a word t to a document in the corpus called D. The input for TF-IDF is obtained by cleaning and extracting the text column in *notesevents* table. Since we use a retrospective topic model, all notes that are written during a patient stay in ICU are combined into one document. The special character, trailing white space, and new line character are removed from clinical notes. A list of stop words is created from combining the Onix stopword list[15] and the default list provided by machine learning library in Spark. All stop words are removed from clinical notes. The LDA parameters including document concentration (α) and topic concentration (β) are set using the following formulas[5].

$$\alpha = \frac{50}{\text{number of topics}} \quad (3)$$

$$\beta = \frac{200}{\text{vocabulary size}} \quad (4)$$

The output of LDA is used as latent topic features. For the mortality prediction, a variety of machine learning algorithms, such as Logistic Regression (LR), Linear Support Vector Machines (SVM), Random Forest (RF), and Gradient Boosted Trees (GBT) are deployed in this project. The target feature or label column are extracted from *hospital_expire_flag* field in the *patient* table. The classifier parameters are tuned using five-fold cross validation on the training data.

The data wrangling, feature engineering, and faceting are performed using OpenRefine, Python, and Scala. Severity scores, MIMIC tables, intermediate views, and queries are built using Spark SQL. Apache Spark is a fast and general engine for big data processing that helps handling the data in this project. Several queries that are available in MIMIC Code Repository[16] are implemented in Spark SQL. A significant number of modifications and customization are made to these queries in order to successfully run them in Spark. The implementation of TF-IDF metrics, topic modeling algorithm (LDA), and classifiers relies on the ML package from Spark. The libraries in Spark allow chaining multiple transformers, models, and algorithm to create a pipeline for the final model. Figure 2 shows an overview of the modeling pipeline.

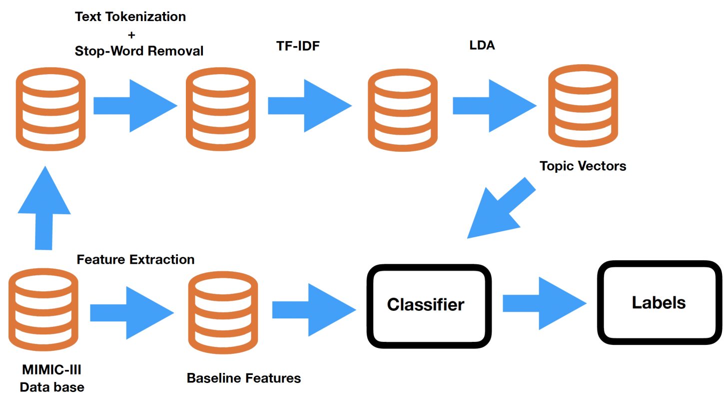

Figure 2: Modeling pipeline

Experimental Results and Discussion

This project uses both AUROC and AUPR to measure comparative performance of the predictive models under various sets of features. The AUROC value that is higher (i.e., closer to 1.0) is preferred since the value represents the probability of ranking a deceased patient higher than alive patients. The value of AUPR depicts a tradeoff between precision and recall (or true positive rate) for various thresholds. A classifier with a higher AUPR returns results with higher recall and precision. In this work, the results are obtained from all valid ICU data. The severity scores are calculated using only the first day of an ICU stay. Since the database contains the records of patients admitted to the hospital or the ICU on more than one occasion, the data is split based on patients ID. This helps to avoid using different records of the same patient in both training and testing data. The data is randomly split into two parts. In this study, 70% of patients are used for training the model and the remaining 30% are used for validation. The classifier's parameters are tuned using five-fold cross validation on the training data. AUROC and AUPR of predictive models are measured on both training and testing data with all classifiers using different sets of features including: structured features, latent topic features, and the combinations of structured and latent topic features. HP Z240 Workstation (Intel Core i7-7700 3.60GHz having 4 physical cores and 8 logical processors with 8MB Cache and 64GB of DDR4 RAM) is deployed for all experiments. The total running time for obtaining the *topics* vector with LDA and the results from training and testing data with all four classifiers is about 8 hours on this platform.

In the pre-processing of inputs for LDA, the minimum number of different documents that a particular term must appear in order to be considered as inputs is set to 10, which derived a vocabulary size of 255,536. In the setting of LDA, the maximum numbers of iterations and topics are set to 200 and 50, respectively. These values are substituted in formula (3) and (4), to set document concentration (α) and topic concentration (β) values. The most informative words from each of the 50 topics are then extracted from the outputs of LDA. As the LDA was being trained and the topics being extracted, more insights are gained which helps to improve the model. For example, initially *Javascript*, *popup*, and *web tag* were among the most relevant words of a topic extracted from the clinical notes. Examining the records shows these words are likely mixed with the clinical notes when the data was extracted from the hospital database. By filtering out these irrelevant words from the input data, LDA produces more meaningful topics. Since there are many medical terms and abbreviations associated with each topic, it is not easy to interpret all of the derived clusters. The first 10 terms associated with the first 50 topics and inferred subjects are listed in a table in appendix A. An analysis of the topics and related words shows there are additional information that can also provide new insights into essential services in the ICU. Specifically, there are records containing information about the the nationality of patients and whether or not they can follow the instructions. For example, 3,877 clinical notes came from 627 distinct patients whose primary spoken language is Russian and some of them need translator or interpreter.

The parameter values for all deployed classifiers (LR, RF, linear SVM, and GBT) are shown in Table 1.

Table 1: Classifiers' parameters

Classifier Name	Parameters
LR	maxIter = 100 , regParam =0.0 , convergenceTolerance = 10^{-6}, Fit Intercept = True
RF	maxDepth = 5, numberOfTree = 400, impurityMeasure = Gini Index
SVM	maxIter = 100, type = Linear kernel , regParam = 0.0, convergenceTolerance = 10^{-6}, Fit Intercept = True
GBT	maxDepth = 5, lossType= logistic , impurityMeasure = Gini Index , stepSize = 0.1

The training results in Table 2 show the knowledge obtained from clinical notes can successfully improve a classifier that only includes structured features. Specifically, when both structured and latent topic features are used for prediction, AUROC achieves its highest value. The best AUROC result on the training data is obtained when GBT is deployed as a classifier using combined (structured and latent topic) features. Comparing the AUPR values obtained from all classifiers also demonstrates that a combined set of features predicts the labels with higher precision and recall values. The best AUPR value is achieved when using the GBT as a classifier with combined features.

Table 2: AUROCs and AUPRs obtained from training with each set of features

Classifier - metric	Structured Feature	Latent Topic Features	Combined Features
LR-AUROC	0.826	0.832	0.869
LR-AUPR	0.733	0.75	0.805
RF-AUROC	0.838	0.826	0.869
RF-AUPR	0.756	0.74	0.807
SVM-AUROC	0.826	0.825	0.868
SVM-AUPR	0.717	0.707	0.78
GBT-AUROC	0.858	0.863	0.903
GBT-AUPR	0.789	0.8	0.863

The performance measures gathered from the testing data are shown in Table 3. The AUROC and AUPR values from the testing data are consistent with the results obtained from the training data. Specifically, each classifier produces the best results when both structured and latent topic features are deployed. The GBT classifier achieves the highest AUROC and AUPR values on the testing data.

Table 3: AUROCs and AUPRs obtained from testing the trained model with each set of features

Classifier - metric	Structured Feature	Latent Topic Features	Combined Features
LR-AUROC	0.831	0.83	0.871
LR-AUPR	0.772	0.771	0.826
RF-AUROC	0.831	0.812	0.86
RF-AUPR	0.77	0.74	0.811
SVM-AUROC	0.826	0.827	0.871
SVM-AUPR	0.744	0.741	0.803
GBT-AUROC	0.841	0.824	0.875
GBT-AUPR	0.781	0.758	0.829

Figure 3 shows the ROC curves derived from the testing data using the GBT classifier with structured, latent topic and combined features. The AUROC values from structured, latent, and combined features are 0.841, 0.811, and 0.874, respectively.

Comparative performance evaluation between the training and testing phases for all classifiers shows a significant improvement of the overfitting problem encountered in the earlier phase of the project when only a subset of input

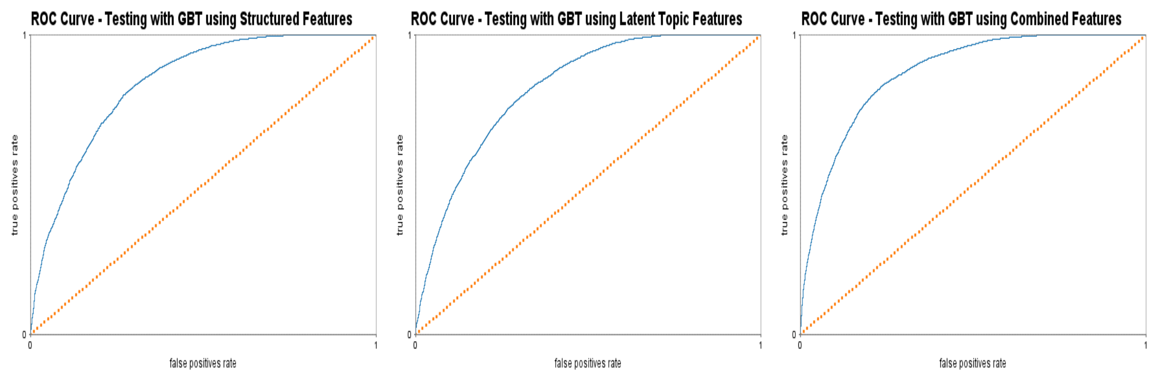

Figure 3: ROC curves obtained from GBT with the structured, latent topic and combined features

records was deployed. A chart depicting the comparative performance of all classifiers for various groups of features on the testing data is shown in Figure 4. As the results show, the GBT classifier achieves the highest AUROC and AUPR on the testing data. The GBT classifer builds a prediction model using an ensemble of decision trees. The best split is chosen based on *Gini Index*. Since the decision trees are deployed as weak learners, we limit the depth of each tree to 5. Gradient descent procedure in the GBT classifier helps adding parameterized trees that optimize the loss function.

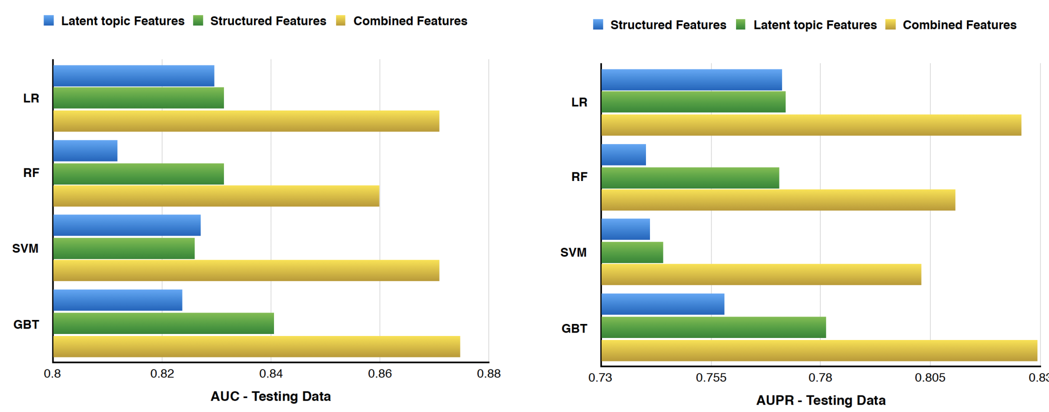

Figure 4: Comparative performance of classifiers using various type of features

Since the selected population and structured features used in this study are not exactly similar to those used in the seminal paper[5], a direct comparison of values is neither logical nor relevant. Nevertheless the results of this study from the four classifiers (LR, RF, Linear SVM, and GBT) show a similar trend as to the results from a sole Linear SVM classifier obtained by Ghassemi et al.[5] where the preditive performance of the classifier is improved when a combination of structured and latent topic features is deployed.

Conclusion

Processing and transforming voluminous electronic healthcare records into accurate and applicable predictive information is a challenging and important endeavor. Significant work on cleaning, wrangling, and feature engineering are required. The usage of ML algorithms and Big Data tools for implementing mortality prediction model also requires significant effort in fine-tuning the parameters. The deployment of a widely adopted database called MIMIC-III also requires significant computing power and memory capacity. This study focused on building and evaluating a mortality prediction model that utilizes both structured and latent topic features derived from the MIMIC-III database. The results demonstrate the latent topic features extracted from the free-text clinical notes help improving the performance of mortality prediction model. Another insight from this study is the extraction of a meaningful set of topics using LDA is a cyclical process. Specifically, analyzing the generated topics and related terms helps filtering irrelevant terms and fine tuning parameters further. Repeating this process helps to improve the final AUROC and AUPR of models that include the latent topic features. The overfitting problem encountered in the earlier phase of the project decreased significantly as more records are deployed and further adjustments are applied to each classifier. A possible direction for improving the LDA outputs is transforming various expressions of the same non-medical or medical term into one expression to avoid repetitive words in the clusters. This study can also be expanded to include the ICU data from other hospitals to build a more generalized model in addition to data from the MIMIC-III database.

References

1. Kim MM, Barnato AE, Angus DC, Fleisher LF, Kahn JM. The effect of multidisciplinary care teams on intensive care unit mortality. Archives of internal medicine. 2010;170(4):369-376. doi:10.1001/archinternmed.2009.521.
2. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der LAAN MJ. Mortality prediction in the ICU: can we do better? Results from the Super ICU Learner Algorithm (SICULA) project, a population-based study. The Lancet Respiratory medicine. 2015;3(1):42-52. doi:10.1016/S2213-2600(14)70239-5.
3. Pirracchio, R. Mortality prediction in the ICU based on MIMIC-II results from the Super ICU Learner Algorithm (SICULA) project. Secondary Analysis of Electronic Health Records. 2016;295-313. doi:10.1007/978-3-319-43742-2_20.
4. Siontis GCM, Tzoulaki I, Ioannidis JP. Predicting death: an empirical evaluation of predictive tools for mortality. Arch Intern Med. 2011;171(19):1721?1726. doi:10.1001/archinternmed.2011.334.
5. Ghassemi M, Naumann T, Doshi-Velez F, Brimmer N, Joshi R, Rumshisky A, Szolovits P. Unfolding physiological state: mortality modeling in Intensive Care Units. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014;75-84. 10.1145/2623330.2623742.
6. Ghassemi M, Naumann T, Joshi R, Rumshisky A. Topic models for mortality modeling in intensive care units. ICML machine learning for clinical data analysis workshop, 2012.
7. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language Processing. Journal of the American Medical Informatics Association. 2004;11(5):392-402. doi:10.1197/jamia.M1552.
8. Solt I, Tikk D, Gál V, Kardkovács ZT. Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier. Journal of the American Medical Informatics Association. 2009;16(4):580-584. doi:10.1197/jamia.M3087.
9. Saria S, McElvain G, Rajani AK, Penn AA, Koller DL. Combining structured and free-text data for automatic coding of patient outcomes. AMIA Annual Symposium Proceedings. 2010;2010:712-716.
10. Arnold C, Speier W. A topic model of clinical reports. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. 2012;1031-1032. DOI: https://doi.org/10.1145/2348283.2348454.
11. Lehman LW, Saeed M, Long W, Lee J, Mark R. Risk stratification of ICU patients using topic models inferred from unstructured progress notes. AMIA Annual Symposium Proceedings. 2012;2012:505-511.
12. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, A freely accessible critical care database. Scientific Data. 2016;3:160035. doi:10.1038/sdata.2016.35.
13. Asuncion A, Welling M, Smyth P, Teh YW. On smoothing and inference for topic models. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. 2009; 27-34.
14. Salton G, Yang CS. On the specification of term values in automatic indexing. Journal of Documentation. 1973. 29(4):351-372.
15. Lextek International-Onix. Onix Text Retrieval Toolkit. 2018. http://www.lextek.com/manuals/onix.
16. MIT Laboratory for Computational Physiology. MIMIC Code Repository: MIMIC-III database. 2018. https://mimic.physionet.org.

Table 4: Appendix A - Examples of derived topics and related words

	Top 10 Related Words	Inferred Subject
1	transplant, seizure, eeg, liver, seizures, keppra, epilepticus, dilantin, dl, renal	Seizure
2	cva, stroke, mri, cerebral, cerebellar, infarction, weakness, aphasia, sided, infarct	Stroke
3	bmt, gvhd, lymphoma, acyclovir, neutropenic, aml, voriconazole, neutropenia, myeloma, bactrim	Blood cancer
4	sdh, subdural, evacuation, craniotomy, hemorrhage, frontal, crani, dilantin, neurosurgery, head	Brain bleeding
5	fracture, fx, trauma, ortho, rib, fractures, fall, hip, spine, injuries	Injury and fracture
6	bipap, niv, hypercarbia, mask, gangrene, mental, altered, hypercarbic, imipenem, chf	Lung disease
7	etoh, ciwa, abuse, valium, withdrawal, alcohol, delirium, haldol, thiamine, ativan	Alcohol abuse
8	cirrhosis, lactulose, liver, encephalopathy, ascites, varices, hepatic, paracentesis, hiv, octreotide	Liver diseases
9	zoster, lupus, italian, derm, dermatology, lyme, naproxen, rash, pmr, bullous	Allergy
10	rvad, tkr, flows, actonel, risedronate, vad, lvad, rv, ami, benzocaine	Heart disease, Bone disease
11	vasculitis, shoulder, anca, hydroxychloroquine, mcp, childhood, joint, raynaud, rml, sarcoidosis, sle	
12	dl, action, assessed, history, mg, assessment, meq, patient, medications, acute	Patient assessment
13	xrt, ent, mass, squamous, carcinoma, decadron, neoplasm, pheresis, stridor, myasthenia	Cancer
14	lap, ileostomy, colectomy, jp, dilaudid, abscess, tpn, ostomy, abdominal, resection	Abdominal surgery
15	pn, il, infant, cbg, fio, diuril, hfov, settings, secretions, simv	
16	hd, dialysis, crrt, esrd, wound, renal, cvvh, failure, stage, cvvhd	Kidney failure
17	ercp, pancreatitis, cholangitis, biliary, cholecystitis, stone, ruq, cbd, gallstone, stent	Pancreas, Liver disease
18	cardiomypathy, extending, ocular, midbrain, clozapine, neuroleptic, guardian, dementia, mal, catastrophic	Neuron imbalance,Heart disease
19	infant, cares, feeds, mom, spells, spits, caffeine, isolette, wt, voiding	Neonatal ICU
20	avr, aortic, cabg, svg, iabp, graft, ci, bypass, coronary, cvicu	Heart disease
21	gsw, melanoma, fasciotomies, rhabdomyolysis, washout, methadone, rhabdo, trauma, inflicted, injury	Surgical treatment
22	aztreonam, obstipation, anaphylaxis, fos, hypercarbic, trach, tamiflu, pyrexia, rash, cephalosporins	
23	klebsiella, parkinson, acinetobacter, tobramycin, hypothyroidism, pseudomonas, trach, amikacin, esbl, meropenem	
24	xigris, cvid, clarithromycin, pancolitis, cisatracurium, ivig, biaxin, aseptic, motrin, ibd	
25	sulfa, sulfonamides, sulfonamide, shellfish, derived, trach, ip, tetracycline, rash, boyfriend	Allergy
26	iodine, hives, hcl, compazine, flolan, desensitization, injection, hickman, containing, demerol	
27	bleed, gib, bleeding, egd, brbpr, melena, hct, colonoscopy, prbc, gastrointestinal	Gastrointestinal bleeding
28	overdose, suicide, psych, narcan, lithium, nac, ingestion, toxicity, sitter, toxicology	Drug overdose
30	tracheal, reconstruction, bronchus, flap, stent, bronch, stenosis, bronchoscopy, tracheobronchial, mainstem	Lung disease
31	dl, assessment, action, cmh, comments, trach, medications, meq, response, pm	
32	osa, obesity, asthma, bipap, hypoventilation, morbid, obstructive, apnea, unknown, morbidly	Breathing problems
33	anorexia, topamax, laminectomy, ocd, lexapro, radiculopathy, rulobectomy, topiramate, clonazepam,lyrica	
34	colitis, diff, difficile, flagyl, cdiff, afib, dl, unknown, clostridium, shock	Colon disease
35	ent, trach, thick, secretions, pt, peep, coarse, gtt, remains, suctioned	
36	hypercapneic, tablet, depakote, divalproex, niece, rectocolectomy, acalc, sessions, prolapse	
37	dl, failure, cmh, assessed, assessment, action, comments, pm, ards, acute	Patient assesment
38	ild, angioedema, ipf, vats, biopsy, steroids, myopathy, progressive, rheum, interstitial	
39	icp, sah, hemorrhage, aneurysm, evd, subarachnoid, mannitol, headache, drain, nicardipine	Brain disease
40	pea, hemoptysis, schizophrenia, arctic, ecmo, tandem, tandemheart, arrest, atropine, poba, cardiogenic, risperidone, ptosis	
41	stent, tbm, tracheobronchomalacia, thoracotomy, bronch, tracheomalacia, rigid, epidural, ip, bronchomalacia	Airway disease
42	arrest, vf, anoxic, sun, cooling, vfib, eeg, cpr, artic, brain	
43	interpreter, pontine, russian, language, hyperthyroidism, deaf, snoring, translation, deafness, english	Patient's language
44	pancreatic, mass, whipple, stricture, pancreas, ampullary, inguinal, unresectable, asbestosis, rcc	Pancreatic disease
45	icd, ep, ablation, vt, paced, amiodarone, ccu, aicd, lidocaine, cardiomyopathy	Abnormal heartbeat
46	futter, aflutter, flut, atrial, amox, clavulanate, augmentin, hida, segmental, heaves, straight, tessalon, copd, wretching	
47	varian, subtotal, adenocarcinoma, uterine, gyn, fasciotomy, endometrial, gj, compartment, vaginal	
48	abetalol, labetolol, hypertensive, sle, malignant, crisis, urgency, emergency, clonidine, hypertension	Blood presuure
49	urology, bladder, prostate, hematuria, nephrostomy, urosepsis, cbi, suprapubic, hydronephrosis, ureteral	Urinary infection
50	keflex, cephalexin, monohydrate, diverticulum, polycythemia, receptor, distension, myelofibrosis, impaction, zenker	