Classification of the Disposition of Patients Hospitalized with COVID-19: Reading Discharge Summaries using Natural Language Processing

Marta Fernandes, Haoqi Sun, Aayushee Jain, Haitham S. Alabsi, Laura N. Brenner, Elissa Ye, Wendong Ge, Sarah I. Collens, Michael Leone, Sudeshna Das, Gregory K. Robbins, Shibani S. Mukerji, M. Brandon Westover

Submitted to: JMIR Medical Informatics
on: November 02, 2020

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Classification of the Disposition of Patients Hospitalized with COVID-19: Reading Discharge Summaries using Natural Language Processing

Marta Fernandes[1, 2, 3*] PhD; Haoqi Sun[2, 3, 1*] PhD; Aayushee Jain[2, 1*]; Haitham S. Alabsi[3, 1]; Laura N. Brenner[3, 4, 5] MD; Elissa Ye[2, 1]; Wendong Ge[2, 3, 1] PhD; Sarah I. Collens[1]; Michael Leone[1]; Sudeshna Das[3, 1] PhD; Gregory K. Robbins[3, 6*] MD; Shibani S. Mukerji[3, 1*] MD; M. Brandon Westover[2, 3, 1, 7*] MD, PhD

[1]Department of Neurology, Massachusetts General Hospital (MGH) Boston US
[2]Clinical Data Animation Center (CDAC) Boston US
[3]Harvard Medical School Boston US
[4]Division of Pulmonary and Critical Care Medicine, MGH Boston US
[5]Division of General Internal Medicine, MGH Boston US
[6]Division of Infectious Diseases, MGH Boston US
[7]McCance Center for Brain Health, MGH Boston US
[*]these authors contributed equally

Corresponding Author:
Marta Fernandes PhD
Department of Neurology, Massachusetts General Hospital (MGH)
50 Staniford St
Boston
US

Abstract

Background: Medical notes are a rich source of patient data, however the nature of unstructured text has largely precluded using these data in large retrospective analyses. Transforming clinical text into structured data can enable large-scale research studies with electronic health records (EHR) data. Natural language processing (NLP) can be used for text information retrieval, reducing the need for labor intensive chart review. Here we present an application of NLP to large-scale analysis of medical records at two large hospitals for patients hospitalized with COVID-19 infections.

Objective: Our study goal was to develop an NLP pipeline to classify the discharge disposition (home, inpatient rehabilitation, skilled inpatient nursing facility (SNIF) and death) of patients hospitalized with COVID-19 based on hospital discharge summaries notes.

Methods: Text mining and feature engineering were applied to unstructured text from hospital discharge summaries. The study included patients with COVID-19 discharged from 2 hospitals in the Boston, Massachusetts area (Massachusetts General Hospital and Brigham and Women’s Hospital) between March 10, 2020, and June 30, 2020. The data was divided into 70% for training and 30% for a hold-out test set. Discharge summaries were represented as bags-of-words consisting of single words (1-grams), 2-grams and 3-grams. The number of features was reduced during training by excluding n-grams that occurred in fewer than 10% of discharge summaries, and further using LASSO regularization while training a multiclass logistic regression model. Model performance was evaluated in the hold-out test set.

Results: The study cohort comprised 1737 adult patients (median [SD] age, 61[18] years old; 55% men; 45% White and 16% Black; 14% non-survivors; 61% discharged home). The model selected 179 from a vocabulary of 1056 engineered features, consisting of combinations of unigrams, bigrams and trigrams. The top features contributing most to the classification by the model (for each outcome) were: ‘appointments specialty', ‘home health’ and ‘home care' (home), 'intubate’, and ‘ARDS’ (inpatient rehabilitation), ‘service’ (SNIF), ‘brief assessment' and ‘covid' (death). The model achieved micro average area under the receiver operating characteristic and average precision in the testing set of 0.98 (95% CI 0.97-0.98) and 0.81 (95% CI 0.75-0.84), respectively, for prediction of discharge disposition.

Conclusions: A supervised learning-based NLP approach is able to classify discharge disposition of patients hospitalized with COVID-19 infection. This approach has the potential to accelerate and increase the scale of research on patients’ discharge disposition that is possible with EHR data. Clinical Trial: Not clinical trial.

(JMIR Preprints 02/11/2020:25457)
DOI: https://doi.org/10.2196/preprints.25457

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Classification of the Disposition of Patients Hospitalized with COVID-19: Reading Discharge Summaries using Natural Language Processing

[1,2,3]Marta Bento Fernandes, PhD, [1,2,3]Haoqi Sun*, [1,3]Aayushee Jain*, [1,2]Haitham S. Alabsi, [2,4,5]Laura N. Brenner,[1,3]Elissa Ye, [1,2,3]Wendong Ge, [1]Sarah I. Collens, [1]Michael Leone, [1,2]Sudeshna Das, [2,6]Gregory K. Robbins**, [1,2]Shibani S. Mukerji**, [1,2,3,7]M. Brandon Westover**

[1]Department of Neurology, Massachusetts General Hospital (MGH), Boston, MA
[2]Harvard Medical School, Boston, MA
[3]Clinical Data Animation Center (CDAC), MGH, Boston, MA
[4]Division of Pulmonary and Critical Care Medicine, MGH, Boston, MA
[5]Division of General Internal Medicine, MGH, Boston, MA
[6]Division of Infectious Diseases, MGH, Boston, MA
[7]McCance Center for Brain Health, MGH, Boston, MA
* Co-first authors
** Co-senior authors

Abstract

Background:
Medical notes are a rich source of patient data, however the nature of unstructured text has largely precluded using these data in large retrospective analyses. Transforming clinical text into structured data can enable large-scale research studies with electronic health records (EHR) data. Natural language processing (NLP) can be used for text information retrieval, reducing the need for labor intensive chart review. Here we present an application of NLP to large-scale analysis of medical records at two large hospitals for patients hospitalized with COVID-19 infections.

Objective:
Our study goal was to develop an NLP pipeline to classify the discharge disposition (home, inpatient rehabilitation, skilled inpatient nursing facility (SNIF) and death) of patients hospitalized with COVID-19 based on hospital discharge summaries notes.

Methods:
Text mining and feature engineering were applied to unstructured text from hospital discharge summaries. The study included patients with COVID-19 discharged from 2 hospitals in the Boston, Massachusetts area (Massachusetts General Hospital and Brigham and Women's Hospital) between March 10, 2020, and June 30, 2020. The data was divided into 70% for training and 30% for a hold-out test set. Discharge summaries were represented as bags-of-words consisting of single words (1-grams), 2-grams and 3-grams. The number of features was reduced during training by excluding n-grams that occurred in fewer than 10% of discharge summaries, and further using LASSO regularization while training a multiclass logistic regression model. Model performance was evaluated in the hold-out test set.

Results:
The study cohort comprised 1737 adult patients (median [SD] age, 61[18] years old; 55% men; 45% White and 16% Black; 14% non-survivors; 61% discharged home). The model selected 179 from a vocabulary of 1056 engineered features, consisting of combinations of unigrams, bigrams and trigrams. The top features contributing most to the classification by the model (for each outcome) were: ‘appointments specialty', ‘home health’ and ‘home care' (home), 'intubate’, and ‘ARDS’ (inpatient rehabilitation), ‘service’ (SNIF), ‘brief assessment' and ‘covid' (death). The model achieved micro average area under the receiver operating characteristic and average precision in the testing set of 0.98 (95% CI 0.97-0.98) and 0.81 (95% CI 0.75-0.84), respectively, for prediction of discharge disposition.

Conclusions:
A supervised learning-based NLP approach is able to classify discharge disposition of patients hospitalized with COVID-19 infection. This approach has the potential to accelerate and increase the scale of research on patients’ discharge disposition that is possible with EHR data.

Keywords: ICU; Coronavirus; Electronic health records; Unstructured Text; Natural Language Processing; BoW; LASSO; Feature Selection; Machine Learning.

Introduction

The COVID-19 pandemic continues to present challenges for healthcare systems around the world [1–8], with over 32.7 million COVID-19 cases confirmed and 991,000 deaths worldwide as of September 27, 2020 [6]. The SARS-CoV-2 virus appeared first in Wuhan, China in December 2019. The first case in the United States (US) was confirmed January 20 [9], followed by rapid spread [2]. By the end of April, Massachusetts became the third hardest hit state, trailing New York and New Jersey [10].

To prepare for a possible second wave in Massachusetts, we set out to conduct a large-scale study of factors associated with outcomes in hospitalized patients at two large academic Boston Hospitals. This effort required the Herculean task of reviewing medical records for over 1000 patients. For structured parts of the electronic health record (EHR), automated data extraction is straightforward. However, some essential information is exclusively or most reliably available only in semi-structured or unstructured narrative medical notes, including patient-reported symptoms, examination findings, or social habits. Thus, developing automated approaches to EHR information extraction wherever possible is critical for more complete patient phenotyping.

Natural language processing (NLP) deals with automated analysis of unstructured text data. Recent advances in NLP machine learning have empowered computers to do several tasks such as machine translation, speech recognition, speech synthesis, semantic understanding and text summarization [11,12]. NLP presents the advantage of being much faster than human chart review of medical records [13–16].

Here we present an automated approach, using NLP, to extract a specific outcome from hospital discharge summaries: discharge destination or “disposition”, i.e. anticipated location or status following discharge. Dispositions of interest included home, inpatient rehabilitation center, skilled inpatient nursing facility, and death. Discharge disposition of patients with COVID-19 from healthcare facilities is important due to the high risk of transmission of the disease within nursing homes and hospitals when patients are discharged to locations other than home, and because it represents an important measure closely related to functional outcome and level of disability following hospitalization, and overall costs of care. Furthermore, this information has the potential to aid healthcare facilities in resource planning to better prepare for the incoming flow of patients. While our model is tailored for discharge disposition, the approach we developed is generalizable to other outcomes available in discharge summaries.

Methods

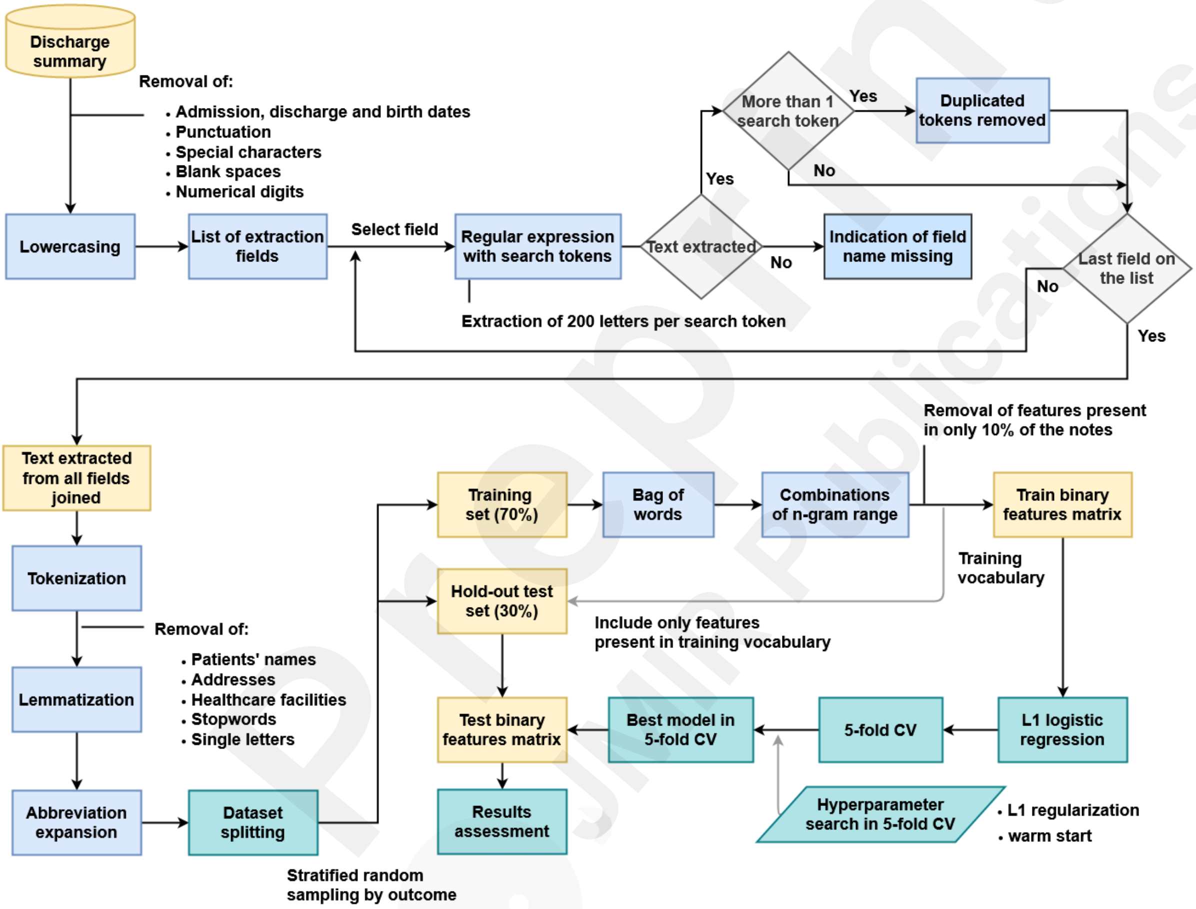
Study Overview

Data was extracted from the hospital electronic medical record under a research protocol approved for a waiver of informed consent by the Partners Healthcare Institutional Review Board. Clinical data were retrospectively analyzed for all adult patients who tested positive for SARS-CoV-2 infection between March 10, 2020 to June 30, 2020. A total of 1737 patients admitted to two major Boston hospitals, 1232 from Massachusetts General Hospital (MGH) and 505 from Brigham’s Women Hospital (BWH), were included. Only patients with a physician discharge summary and available known ground-truth discharge disposition were included.

Data Collection and Processing

Data consisted of discharge summaries, which are unstructured free text notes written by physicians, and a ground truth record of discharge disposition, used to assess the accuracy of the NLP results. The methodology for note preprocessing is shown in Figure 1. The upper part of the figure provides an overview of the text extraction for each field on the list of extraction fields depicted in Table 1. The lower part of the figure shows the methodology steps where the text extracted from all the fields is processed for modeling. The data was randomly stratified into train and test sets for modeling, which we address in the Model development section.

Figure 1. Methodology steps for discharge summary notes preprocessing and modeling. The list of extraction fields is depicted in Table 1.



Document preprocessing

Admission, discharge, and birth dates were removed from the discharge summaries, as well as punctuation, special characters, blank spaces and numerical digits. Notes were then subjected to lowercasing, tokenization and correction using lemmatization, a procedure for obtaining the root form of the word, using vocabulary (dictionary importance of words) and morphological analysis (word structure and grammar relations). *WordNetLemmatizer* from NLTK library in Python version 3.7 was used with a POS tag specified as verb. Patients' names, addresses, healthcare facilities and hospital unit names were removed, as well as single letters. Abbreviation expansion and spell corrections were performed for a small list of frequently used clinical words (Table A1, Appendix A). A list of commonly used words less informative *stopwords* was also removed from the notes (Table

A2, Appendix A).

Processing of specific discharge summary fields

Discharge summaries at MGH and BWH are semi-structured, with a series of named fields containing specific types of mostly free text information (Table 1). We present an example of discharge summary notes with protected health information removed (Table A3, Appendix A). Text fields were identified based on information extracted from the notes using regular expressions with search tokens (Table 1). The function 'str.extractall' from Python was used to extract a length of 200 letters of text onwards from all instances where the search token appeared.

Table 1. Information captured from discharge summaries, grouped in fields, and respective search tokens used in the regular expression.

Field	Search token
Discharge disposition	'discharge', 'discharged', 'dispo', 'skilled nursing', 'snf'
Diagnosis	'diagnosis', 'diagnoses', 'problem', 'reason for admission', 'chief complaint'
Surgeries	'surgeries this admission'
Treatments	'treatments'
Tests	'tests'
Allergies	'allergies', 'allergic'
Diet	'diet', 'nutrition'
Medical history	'history'
Hospital course	'hospital course'
Laboratory results	'labs'
Activity	'activity', 'activities'
Physical Exam	'discharge exam', 'physical exam'
Physical therapy	'physical therapy'
Occupational therapy	'occupational therapy'
Discharge instructions	'instructions'
Follow-up care	'follow up'
Discharge plan	'discharge plan'
Additional orders	'additional orders'
Code status	'code status'

Some notes contained a 'discharge disposition' field used to list the discharge disposition. We deleted this field to avoid an overly "easy" solution, because this field is not universally available, and because we wished to assess how well the approach is able to perform when structured data is unavailable. In a field where more than one extraction was performed, i.e. with more than one search token, the corresponding results were joined, and duplicated words were removed. To illustrate with an example, for the 'Diet' field, using the regular expressions with search tokens 'diet' and 'nutrition', 200 letters were captured for each search token, for a total of 400 letters. Since there might be repeated information in the discharge summary regarding diet and nutrition recommendations, duplicated words were removed from the captured text. Where no data was captured with the search tokens, an indication of missingness was set with the name of the field and the suffix '_missing'.

The text extracted from all fields (depicted in Table 1) were joined to create a reduced version of the discharge summary, which was then subjected to tokenization, lemmatization and abbreviation expansion, as described in the Document preprocessing subsection. The vocabulary used for modeling was created based on these reduced versions of the discharge summaries contained in the training set. Each documents were represented as a binary Bag of words (BoW), i.e. an ordered series of binary vectors indicating whether a given n-gram (word or sequence of 2 or 3 words) is present in the document, disregarding grammar and word order. The function *CountVectorizer* was used with its default parameters from Python, except for the n-gram range which was set as unigrams (a single word), bigrams (two consecutive words) and trigrams (three consecutive words). As a first step to reduce dimensionality, only features present in at least 10% of the reduced version of the discharge summary notes were considered. Multi-class logistic regression with the least absolute shrinkage and selection operator (LASSO) [17] was used to further sparsify the model.

Outcome Measure

The multiclass outcome measure was discharge disposition, composed of the classes: home, inpatient rehabilitation, skilled inpatient nursing facility (SNIF) and death. “Home” included “home or self-care”, “home-health care services” and patients who “left against medical advice”. SNIF included “Skilled Nursing Facility” and “Custodial Care Facility”.

Model development

The training algorithm used the one-vs-rest scheme for multiclassification, where a binary problem was fitted for each class and the class weight was balanced. Logistic regression [18] with LASSO regularization was used as the classification model. The model estimator $\hat{\beta}$ is depicted in equation (1) and the LASSO regularization objective can be written as in equation (2). $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\mathbf{Y} \in \mathbb{R}^{n\times 1}$ correspond to the design input matrix and the vector of observations, respectively, where n is the number of observations, in this case number of discharge summaries or number of patients, and p the number of features in $\boldsymbol{x} \in \mathbb{R}^p$. The vector of regression coefficients is given by $\beta \in \mathbb{R}^p$, $\|\beta\|_1$ corresponds to the L1 norm of this coefficients vector and λ is the regularization parameter that controls the amount of shrinkage. The regularization adds a penalty on the weights to prevent overfitting [19]. The inverse of the regularization strength *C* was varied for the values {0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5}.

$$\hat{\boldsymbol{\beta}}(\boldsymbol{x}) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{\beta})}} \text{ (1)}$$

$$\text{minimize } \|\boldsymbol{X}\beta - \boldsymbol{Y}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \text{ (2)}$$

Stratified random sampling was used to split the dataset into a train set (70%) and a hold-out test set

(30%). A randomized search was used for hyperparameter tuning during training with 100 iterations of 5-fold cross validation (CV). The solver was set to “liblinear” and the “warm start” hyperparameter was varied between true/false, where “true” corresponded to reusing the solution of the previous call to fit as initialization, and “false” corresponded to erasing the previous solution.

Performance measures

The R^2 coefficient of determination score was used in CV scoring to select the best model configuration in the training data. The one standard error rule was used to select the regularization parameter. The simplest model, whose R^2 mean score fell within 1 standard deviation of the maximum R^2, was selected.

To measure model performance on test data, the area under the receiver operating characteristic curve (AUROC) was calculated. The ROC curve is a function of recall (sensitivity) versus the false positive rate (FPR; i.e. 1-specificity) (A1). The pair ($Recall_k$, FPR_k) is called an operating point for this curve, where k is a threshold that is varied to generate the ROC curve. The equations for these metrics are presented in Table A4 in Appendix A.

The area under the precision-recall curve (AUPRC) was also calculated, which is an important measure in the presence of class imbalance. The pair ($Recall_k$, $Precision_k$) is referred to as an operating point for this curve. Average precision (AP) (A3) summarizes this plot as the weighted mean of precisions achieved at each threshold, with the increase in recall from the previous threshold used as the weight.

The F1-score (F1) (A4) was also assessed as another performance metric commonly reported for datasets with imbalanced numbers across classes [20].

100 iterations of bootstrap random sampling with replacement was performed to calculate 95 % confidence intervals (CI) for performance metrics.

Results

Summary of patient population

From 1917 patients’ medical records, 1752 had a physician discharge summary and discharge disposition, within the categories of home, inpatient rehabilitation, SNIF and death. Only adults (age ≥ 18 years old) were included in the analysis, leaving a study cohort of 1737 patients. The cohort was split using stratified random sampling according to outcome into train and test sets. Age in the train and test sets was balanced with a median of 62 and 60 years old, respectively (Table 2). The majority of patients were White 774 (median 44.6%) and Black or African American 285 (median 16.4%). Most were discharged home 1052 (60.6%). Among all patients with COVID-19, there were 243 (14.0%) non-survivors.

Table 2. Baseline characteristics of the study patient population stratified by train and test sets.

Characteristic	Train set	Test set	Total

	(n=1215)	**(n=522)**	**(n=1737)**
Age (years old) median (std)	62.0 (18.2)	60.0 (18.2)	61.0 (18.2)
Gender no. (% of total)			
Female	545 (44.9)	244 (46.7)	789 (45.4)
Male	670 (55.1)	278 (53.3)	948 (54.6)
Race no. (% of total)			
White	533 (43.9)	241 (46.2)	774 (44.6)
Hispanic or Latino	52 (4.2)	19 (3.6)	71 (4.1)
Black or African American	204 (16.8)	81 (15.5)	285 (16.4)
Asian	46 (3.8)	21 (4.0)	67 (3.9)
American Indian or Alaska Native	31 (2.5)	13 (2.5)	44 (2.5)
Native Hawaiian or other Pacific Islander	2 (0.2)	1 (0.2)	3 (0.2)
Unknown [a]	347 (28.6)	146 (28.0)	493 (28.3)
Institution no. (% of total)			
MGH	881 (72.5)	351 (67.2)	1232 (70.9)
BWH	334 (27.5)	171 (32.8)	505 (29.1)
Discharge disposition no. (% of total)			
Home	736 (60.6)	316 (60.5)	1052 (60.6)
Inpatient rehabilitation	102 (8.4)	44 (8.4)	146 (8.4)
SNIF [b]	207 (17.0)	89 (17.1)	296 (17.0)
Death	170 (14.0)	73 (14.0)	243 (14.0)

[a] Unknown includes ‘other’, ‘declined’ or ‘unavailable’.
[b] SNIF – skilled inpatient nursing facility.

The preprocessed dataset for modeling was created based on the notes extracted in all fields except the ‘discharge disposition’ and ‘code status’ fields, as described in the 'Processing of specific discharge summary fields’ subsection. Before dimensionality reduction, where features present in at least 10% of the reduced version of the discharge summary notes were considered, there was a total of 15182 tokens (unigrams). After applying this dimensionality reduction step, we were left with 477 tokens. With this set of tokens, 3497 combinations of n-grams were generated, leading to a total of 1056 features with duplicates removal. Thus, the total number of candidate features in the training vocabulary was 1056, including 460 unigrams, 329 bigrams and 267 trigrams.

Modeling Results

The best model configuration parameters and performance results in the hold-out test set are presented in Table 3 with 95% confidence intervals. The corresponding confusion matrices normalized by precision and recall are presented in Figure 2. The performance discriminated by discharge outcome is presented in Table 4. Higher performance was obtained for the outcomes of home discharge and death compared to inpatient rehabilitation and SNIF discharge outcomes. The model presented higher recall (0.95) and precision (1.0) for the death outcome. Home disposition also presented high performance for these metrics. For this model, 2 deceased patients were classified as discharged home. In experiments, for models where we included the discharge disposition field, extracted from the discharge summary, all deceased patients were correctly classified. The inpatient rehabilitation outcome presented the lowest recall (0.61) and 12 patients with this outcome were incorrectly classified by the model as discharged to SNIF. The outcome of disposition to SNIF presented the lowest precision (0.68) overall and 20 patients discharged home were incorrectly predicted as discharged to SNIF. Compared to the initial set of features in the training vocabulary, the final model contained approximately 83% fewer features, with a total of 179 features. The relative importance of the top 30 model features is presented in Figure 3, where the importance for each feature consisted of the sum of the absolute coefficients' values across the outcomes.

Table 3. Model performance in the hold-out test set and configuration parameters. [a]

AUROC	ACC	Recall	F1	AP	Precision	Parameters
0.98 [0.97-0.98]	0.88 [0.85-0.90]	0.88 [0.85-0.90]	0.88 [0.85-0.90]	0.81 [0.75-0.84]	0.88 [0.85-0.90]	No. features (1, 2, 3 grams): 179 (95, 52, 32) C = 0.09 Warm start True

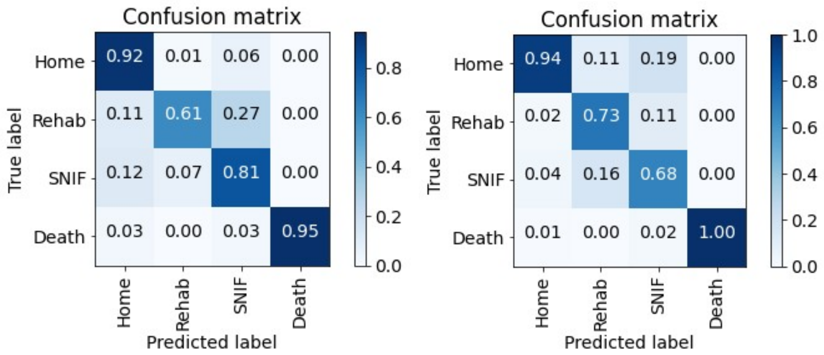
[a] In parenthesis are the bootstrapping results in 95% confidence intervals. AUROC - Area under the receiver operating characteristic curve, ACC - accuracy, AP - average precision.

Table 4. Model performance in the hold-out test set by discharge outcome. [a]

Outcome	AUROC	ACC	Recall	F1	AP	Precision
Home	0.97 [0.95-0.98]	0.92 [0.89-0.94]	0.92 [0.89-0.95]	0.93 [0.91-0.95]	0.92 [0.88-0.94]	0.94 [0.91-0.97]
Rehab	0.95 [0.91-0.98]	0.95 [0.93-0.97]	0.61 [0.53-0.76]	0.67 [0.53-0.78]	0.48 [0.32-0.64]	0.73 [0.58-0.86]
SNIF	0.93 [0.88-0.96]	0.90 [0.87-0.92]	0.81 [0.72-0.88]	0.74 [0.64-0.79]	0.58 [0.46-0.66]	0.68 [0.58-0.75]
Death	1.00 [1.00-1.00]	0.99 [0.99-1.00]	0.95 [0.90-0.98]	0.97 [0.95-0.99]	0.95 [0.91-0.98]	1.00 [1.00-1.00]

[a] In parenthesis are the bootstrapping results in 95% confidence intervals. AUROC - Area under the receiver operating characteristic curve, ACC - accuracy, AP - average precision.

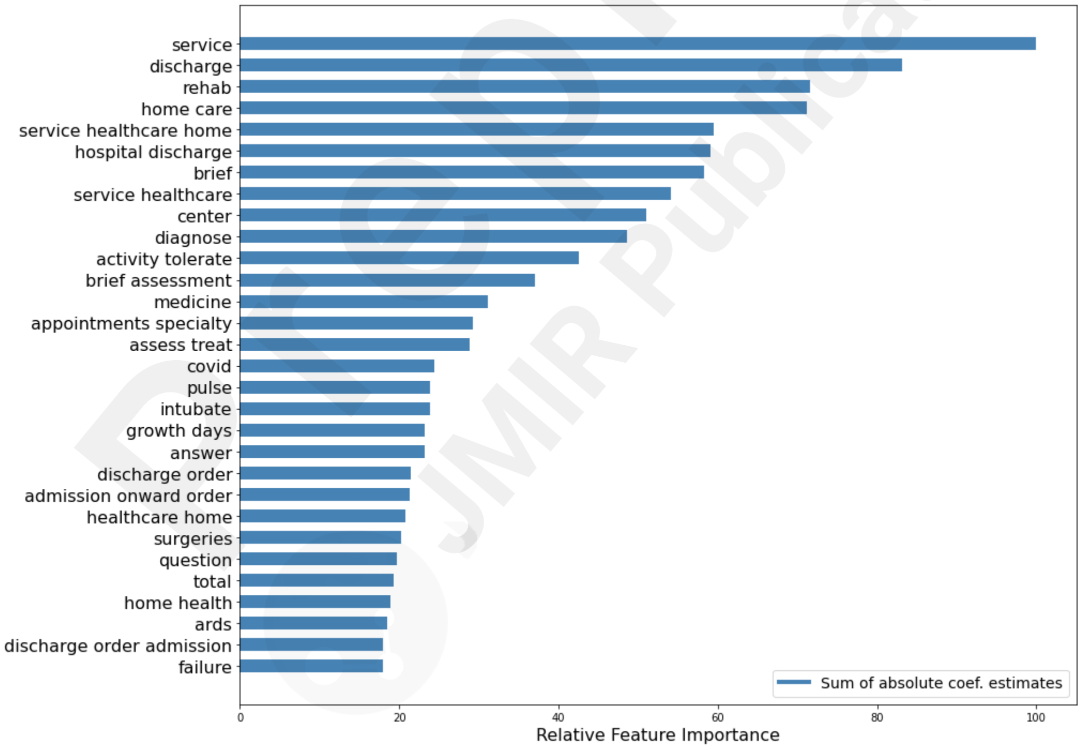
Figure 2. Confusion matrices for the best model evaluated in the hold-out test set normalized (a) by recall and (b) by precision.



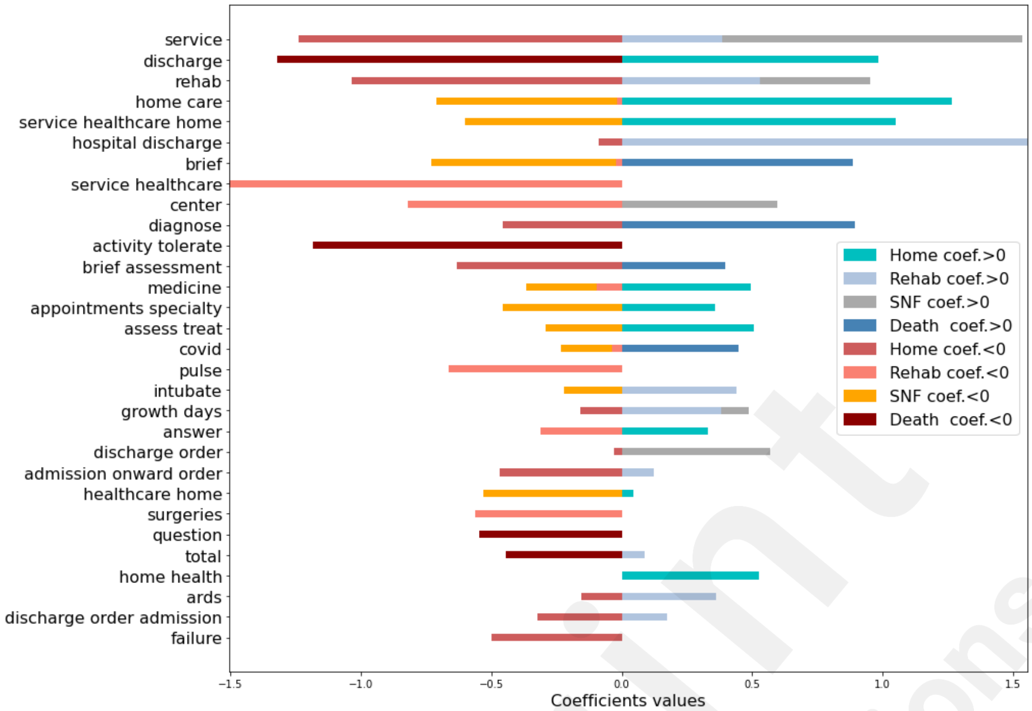
(a) (b)

‘Service’ was the feature assigned the highest importance for classification of the discharge outcomes. For inpatient rehabilitation and SNIF dispositions, the coefficient values for this feature are positive, which indicates that most likely this term will appear in the preprocessed notes for both outcomes. ‘Home care’, ‘healthcare home’ and ‘home health’ were assigned a positive coefficient value for home disposition. ‘Service healthcare home’ was also assigned high importance for this outcome, suggesting that this feature is related to patients discharged home with home-health care services provided. ‘Medicine’ and ‘appointments specialty’ were also important for this outcome. ‘Rehab’ had positive coefficients for both inpatient rehabilitation and SNIF dispositions. ‘Intubate’ and ‘ARDS’ (acute respiratory distress syndrome) are important features for inpatient rehabilitation disposition. For death, ‘discharge’ and ‘activity tolerate’ presented negative coefficients values, indicating that these features are unlikely to appear in discharge summaries of deceased patients. ‘Brief assessment’ and ‘brief’ are assigned high coefficient values for this outcome. ‘Covid’ was assigned a positive coefficient value for predicting death, while for inpatient rehabilitation and SNIF were given negative values.

Figure 3. Relative importance of top 30 features obtained with the model coefficients estimates for (a) the sum of the absolute coefficients values and (b) the coefficients values discriminated by outcome.



(a)



(b)

The training performance is depicted in Appendix B, Figure B1, with the curve correspondent to the R^2scores for the different values of the inversed regularization strength. The top 15 features and their relative importance obtained with LASSO regularization are presented for each outcome (Figure B2). Blue bars correspond to features with positive coefficients values and red bars to features with negative coefficients values. The areas under the ROC and Precision-Recall curves for the best model are also presented (Figure B3). We also assessed how the model performance and the features selected as the most important in train, varied with the dimension of the train set (Figure B4). The hold-out test set for model evaluation was fixed and the train set dimension was varied from 10% to 100% of the original train set, with 1215 patients. We observed that the best performance was achieved with higher number of patients in the train set, the original train set (100%). However, with 50% vs 100% of the original train set, AUROC (0.97 vs 0.98) and AP (0.79 vs 0.81), the model achieved good performance for 1018 (vs 1056) vocabulary features. We assessed the common features between each train set and the original train set (Figure B5). Among the top 30 features, there were 10 common features between the 50% and the original train sets. The higher number of common features was found for the train set with 90% of the original train set, with a total of 17 common features. Finally, we observed that more than half of the features in the top 30, from the original train set, were selected as top 30 at least in two train sets (Figure B6).

Discussion

Principal Findings

In this study a machine learning-based NLP pipeline was developed to classify the discharge disposition of adult patients hospitalized with COVID-19. The model achieved near perfect identification of patients with outcomes of home disposition or death. For intermediate outcomes of inpatient rehabilitation or SNIF, performance was imperfect but also acceptable. Due to this classification task being relatively easy, more complex and time costly modeling approaches, such as recurrent neural networks or bidirectional encoder representations from transformers were not considered. We acknowledge that for harder tasks, these approaches can improve performance. The final method is automated, thus enabling large-scale rapid processing of thousands of discharge summaries, a task that is infeasible when relying on manual chart review.

Limitations

The present analysis was limited to a cohort of patients with COVID-19, who may have specific medical symptoms related to the disease. Therefore, as future work it is proposed to extend the model to diverse other cohorts. Further, although results spanned 2 hospitals, they are located in the same geographic region (Boston, Massachusetts). Thus, our cohort may not be representative of other US and non-US populations. Moreover, decision-making for discharge disposition may vary for different hospitals, according to the number of SNIFs or rehabilitation centers in the geographic area, which may affect the generalizability of the model. The models were developed with textual information from discharge summaries while the addition of other clinical features, such as physical or occupational therapy reports, social work or case manager notes, was not considered, which is a limitation of the study and can be pursued in future work.

Comparison with Prior Work

Extraction of information from clinical narratives is a growing application of NLP in healthcare. NLP has been used to extract information from hospital discharge notes about medical conditions such as postsurgical sepsis [21], pneumonia [22], or other potential medical problems [23], to identify critical illness [24,25], to detect adverse events [26], to predict risk of rehospitalization [27], to extract medication information [28] and to risk stratify patients [29]. To the best of our knowledge, ours is the first work on classifying hospital discharge disposition based on discharge summaries notes using machine learning and NLP.

Conclusions

This study shows that a supervised learning-based NLP approach can be used to accurately classify discharge disposition of hospitalized patients with COVID-19 in an automated fashion. This model, and the NLP approach used to develop it, have the potential to accelerate and increase the scale of research that is possible with EHR data.

Acknowledgements

MBW was supported by the Glenn Foundation for Medical Research and American Federation for Aging Research (Breakthroughs in Gerontology Grant); American Academy of Sleep Medicine (AASM Foundation Strategic Research Award); Football Players Health Study (FPHS) at Harvard University; Department of Defense through a subcontract from Moberg ICU Solutions, Inc; by the

NIH (1R01NS102190, 1R01NS102574, 1R01NS107291, 1RF1AG064312). MBW is a co-founder of Beacon Biosignals.

Conflicts of Interest

There are no conflicts of interest.

Abbreviations

ACC: Accuracy
AP: Average precision
ARDS: acute respiratory distress syndrome
AUPRC: Area under the precision recall curve
AUROC: Area under the receiver operating characteristic curve
BoW: Bag-of-words
BWH: Brigham’s Women Hospital
CI: Confidence interval
COVID-19: Coronavirus disease of 2019
CV: Cross-validation
EHR: Electronic health record
FPR: False positive rate
ICU: Intensive Care Unit
LASSO: Least absolute shrinkage and selection operator
MGH: Massachusetts General Hospital
NLP: Natural Language Processing
PHI: Protected Health Information
PPV: Positive predictive value
ROC: Receiver operating characteristic
SNIF: skilled inpatient nursing facility

References

1. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. Jama. 2020;
2. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. N Engl J Med. 2020;
3. Fauci AS, Lane HC, Redfield RR. Covid-19—navigating the uncharted. Mass Medical Soc; 2020.
4. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The lancet. 2020;
5. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. Jama. 2020;323[13]:1239–1242.
6. COVID TC, Team R. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19)-United States, February 12-March 16, 2020. MMWR Morb Mortal Wkly Rep. 2020;69[12]:343–346.

7. He X, Lau EH, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020;26[5]:672–675.

8. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med. 2020;172[9]:577–582.

9. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. Jama. 2020;323[11]:1061–1069.

10. The Covid-19 Tracker - STAT [Internet]. [cited 2020 Oct 13]. Available from: https://www.statnews.com/feature/coronavirus/covid-19-tracker/

11. Nallapati R, Zhou B, Gulcehre C, Xiang B. Abstractive text summarization using sequence-to-sequence rnns and beyond. ArXiv Prepr ArXiv160206023. 2016;

12. Hirschberg J, Manning CD. Advances in natural language processing. Science. 2015;349[6245]:261–266.

13. Wilbur WJ, Rzhetsky A, Shatkay H. New directions in biomedical text annotation: definitions, guidelines and corpus construction. BMC Bioinformatics. 2006;7[1]:1–10.

14. Buchan NS, Rajpal DK, Webster Y, Alatorre C, Gudivada RC, Zheng C, et al. The role of translational bioinformatics in drug discovery. Drug Discov Today. 2011;16[9–10]:426–434.

15. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc. 2011;18[5]:544–551.

16. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc. 2011;18[5]:552–556.

17. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Methodol. 1996;58[1]:267–288.

18. Cramer JS. The origins of logistic regression. 2002;

19. Bühlmann P, Van De Geer S. Statistics for high-dimensional data: methods, theory and applications. Springer Science & Business Media; 2011.

20. Azari A, Janeja VP, Levin S. Imbalanced learning to predict long stay Emergency Department patients. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2015. p. 807–814.

21. Arvind V, Cho B, Ukogu CO, Kim J, Cho SK. Wednesday, September 26, 2018 2: 00 PM–3: 00 PM Integrating Technology into Practice: 59. Natural language processing of electronic medical records can identify sepsis following orthopedic surgery. Spine J. 2018;18[8]:S29.

22. Yetisgen-Yildiz M, Glavan BJ, Xia F, Vanderwende L, Wurfel MM. Identifying patients with pneumonia from free-text intensive care unit reports. In: Proceedings of Learning from Unstructured Clinical Text Workshop of the International Conference on Machine Learning. 2011.

23. Meystre S, Haug P. Improving the sensitivity of the problem list in an intensive care unit by using natural language processing. In: AMIA Annual Symposium Proceedings. American Medical Informatics Association; 2006. p. 554.

24. Weissman GE, Harhay MO, Lugo RM, Fuchs BD, Halpern SD, Mikkelsen ME. Natural language processing to assess documentation of features of critical illness in discharge documents of acute respiratory distress syndrome survivors. Ann Am Thorac Soc. 2016;13[9]:1538–1545.

25. Marafino BJ, Park M, Davies JM, Thombley R, Luft HS, Sing DC, et al. Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. JAMA Netw Open. 2018;1[8]:e185097–e185097.

26. Murff HJ, Forster AJ, Peterson JF, Fiskio JM, Heiman HL, Bates DW. Electronically screening discharge summaries for adverse medical events. J Am Med Inform Assoc. 2003;10[4]:339–350.

27. Kang Y, Hurdle J. Predictive Model for Risk of 30-Day Rehospitalization Using a Natural Language Processing/Machine Learning Approach Among Medicare Patients with Heart Failure. J

Card Fail. 2020;26[10]:S5.
28. Yang H. Automatic extraction of medication information from medical discharge summaries. J Am Med Inform Assoc. 2010;17[5]:545–548.
29. Lehman L, Saeed M, Long W, Lee J, Mark R. Risk stratification of ICU patients using topic models inferred from unstructured progress notes. In: AMIA annual symposium proceedings. American Medical Informatics Association; 2012. p. 505.

Appendix A

Table A1. List of abbreviations expansion and spell corrections from the lemmatization in the notes preprocessing stage.

Abbreviation, expansion
'abd', 'abdominal' ‘abdo','abdominal' 'afib', 'atrial fibrillation' 'covi', 'covid' 'diagno', 'diagnosis' 'orient', 'orientation' 'tota', 'total' 'disch', 'discharge' 'demonst', 'demonstrate' 'servi', 'services' 'requir', 'require' 'preferen', 'preference' 'preferenc', 'preference' ‘caregive', ‘caregiver'

Table A2. List of stopwords removed from the notes in the preprocessing stage.

Stopwords
'i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', ‘youre’, ‘youve', ‘youll’, ‘youd’, 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "shes", 'her', 'hers', 'herself', 'it', 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'the', 'with', 'to', 'be', 'from', 'which', 'dob', 'date', 'summary', 'please', 'dear', 'fa', 'facesheet', 'wi', 'llst', 'items', 'st', 'same', 'phone', 'sign', 'about', 'should', 'as', 'or', 'an', 'for', 'of', 'ac', 'in', 'by', 'at', 'fo', 'me', 'nan', 'hea', 'pro', 'and', 'up', 'full', 'code', 'have', 'has', 'is', 'part', 'do', 'will', 'there', 'faci', 'this', 'that', 'what', 'nc', 'comment', 'other', 'throughout', 'md', 'mdd', 'qd', 'per', 'sig', 'bid', 'when', 'use', 'while', 'apt', 'resu', 'con', 'dis', 'go', 'doct', 'mch', 'wnl', 'ml', 'mg', 'diff', 'tid', 'id', 'hs', 'medic', 'contact', 'but', 'hid', 'post', 'nt', 'first', ‘if', 'then', ‘who’, ‘whom’, ‘these’, ‘those’, ‘am’, ‘are’, ‘was’, ’were’, ‘been’, ‘being’, 'had', 'having', 'does', 'did', 'doing', 'because', 'until', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'down', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'once', 'here', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'some', 'such', 'nor', 'only', 'own', 'so', 'than', 'too', 'very', 'can', 'just', 'don', 'now'

Table A3. Example of discharge summary notes for a patient where protected health information (PHI) was removed, including the changing of dates.

Discharge summary notes
Physician Discharge Summary Admit date: Discharge date: Patient Information y.o. (DOB = //) Home Address: Home Phone: (home) What language do you prefer to use when discussing your healthcare?: English What language do you prefer for written communication?: English Type of Advance Care Directive(s): None Does patient have a Health Care Proxy form completed?: Patient declines Health Care Agents There are no Health Care Agents on file. Code Status at Discharge: Full Code (Confirmed) Hospitalization Summary Reason for Admission: Fever, Cough, COVID-19 Exposure Principal Problem: 2019 Novel Coronavirus Disease (Covid-19) Resolved Problems: * No resolved hospital problems. * Surgical (OR) Procedures: Surgeries this admission None Procedures this admission None Non (OR) Procedures: Items for Post-Hospitalization Follow-Up: PCP [ ] Please follow up COVID-19 results and resolution of respiratory symptoms. Pending Results Procedure Component Value Ref Range Date/Time COVID-19 RT PCR [] Collected: // Lab Status: In process Updated: // Hospital Course with depression, OSA on CPAP, who has a wife positive for COVID19, presents with cough and fever c/f COVID19 a/f monitoring ----- HPI ----- with depression, OSA on CPAP, presents with progressive cough and fever, and confusion that prompted the wife (cell) to call 911. Per EMS report, patient was back to normal baseline mental status. Patient's wife was transported in the ambulance and since she is know COVID-19 positive, patient was sent home with the ambulance. On both husband and wife presented to the emergency department last week with body aches, myalgia, feeling unwell -Wife tested positive for corona virus, husband tested negative. Husband said he had started to feel "off" from his baseline Approximately 3 days prior to admission, patient started having fevers -Fevers have been coming down with Tylenol and ibuprofen although the fever curve is rising -Fevers to 101, 102 °F -This morning had fever to 101.7 - Upon waking from sleep, was a little bit more confused than normal - Went to the bathroom, was feeling hot, took temperature and found it was 101.7 - Was not wearing CPAP overnight as usually does - patient was found to be confused by wife - wife reported patient has been sleeping longer than usually Upon arrival to the ED, patient was satting well on RA, in no distress. He became febrile to 100.5F, otherwise vitals have been stable. He was given Tylenol and tessalon pearls. Labs s/f negative procal, normal WBC, and negative respiratory viral panel. SARS-COV2 swab was sent and is pending. CXR showed "Equivocal faint hazy upper lung opacities bilaterally". Upon my interview, patient said he was feeling fine. +dry cough, congestion, poor appetite. He had some diarrhea earlier the day prior which he thought was from ibuprofen but hadn't had any since. No SOB, CP, N/V, HA, dizziness. ----- SPU Course ----- #C/f COVID19 #Fever, cough Patient presents with 3 days of symptoms. Fever, fatigue, dry cough, anorexia. His wife is positive for COVID19. Differential diagnosis includes COVID vs other viral illness. Forunately, no leukopenia of LFT abnormalities. Has negative flu/ RSV/ Adeno/ Human metapneumovirus PCR/ Rhinovirus/ Parainfluenza/ Influenza A, B, RSV negative. Not likely to be bacterial, negative procal (0.10). Respiratory status stable. Patient had one episode of fever to 100.5 F during admission but defervesced with Tylenol and has been afebrile since. Very low risk for decompensation given age and no medical comorbidities. Patient was admitted for COVID-19 rule out given episode of confusion at home and CXR finding of equivocal faint hazy upper long opacities. Repeat CXR was stable. Patient's cough was managed with tessalon and

robitussin PRN. At the time of discharge patient's COVID-19 RT PCR was pending. Given clinical stability, patient was discharged to home with instructions for home isolation. Boston Department of Health was called and informed of patient's case. Patient was instructed to self isolate and given strict instructions regarding return precautions. #Depression C/h wellbutrin 300mg daily Medications Allergies: Patient has no known allergies. Prior to Admission Medications Prescriptions buPROPion (WELLBUTRIN XL) 300 MG ER 24 hr tablet Sig: Take 300 mg by mouth daily. Facility-Administered Medications: None Medication List TAKE these medications Instructions benzonatate 100 MG capsule Commonly known as: TESSALON Last time this was given: Take 1 capsule (100 mg total) by mouth 3 (three) times a day as needed for cough. buPROPion 300 MG ER 24 hr tablet Commonly known as: WELLBUTRIN XL Take 300 mg by mouth daily. ondansetron 4 MG tablet Commonly known as: ZOFRAN Take 1 tablet (4 mg total) by mouth every 12 (twelve) hours as needed for nausea. Where to Get Your Medications These medications were sent to CVS/pharmacy Phone benzonatate 100 MG capsule ondansetron 4 MG tablet Hospital Care Team Service: Medicine Inpatient Attending: Attending phys phone: Discharge Unit: Primary Care Physician: Not Required Pcp None Transitional Plan Scheduled appointments: Signed Discharge Orders (From admission, onward) Ordered // Activity as tolerated // Discharge diet Comments: Diet Regular // For immediate questions regarding your hospitalization, your medications, and any pending test results please contact your doctor in the hospital: MD. Comments: For immediate questions regarding your hospitalization, your medications, and any pending test results please contact your doctor in the hospital: MD. Discharge instructions and important events and results You were admitted to BWH on // with cough and fever. Given your recent exposure to COVID-19, you were admitted to the Special Pathogens Service and tested for this virus. These results were still tested at the time of your discharge. Given the stability in your respiratory symptoms, you were discharged with a plan to self-quarantine at home until your results return. You will be contacted regarding the results by your local department of health and the best next steps after receiving these results. You can take Tylenol 650 mg every 8 hours to treat your fevers. Your last dose of 975 mg was at 12:40 PM today. You can start taking Tylenol 650 mg every 8 hours starting at 6 PM this evening. You have been discharged home with new medications: - Tessalon Perles 100 mg three times daily as needed for cough. - Zofran 4 mg twice daily as needed for nausea. Please contact us with any questions or concerns. It was a pleasure taking care of you and we wish you a speedy recovery! Sincerely, The BWH team Exam Temperature: 37.7 °C (99.8 °F) (03/08/20 1532) | Heart Rate: 95 (03/08/20 1532) | BP: 121/77 (03/08/20 1532) | Respiratory Rate: 16 (03/08/20 1225) | SpO2: 100 % (03/08/20 1532) O2 Device: None (Room air) (03/08/20 1532) | Weight: 97.5 kg (215 lb) (03/08/20 2315) Height: 172.7 cm (5' 8") (03/07/20 2315) BMI (Calculated): 32.7 (03/07/20 2315) Discharge Exam Significant Discharge Exam Findings: General: Well-developed, no apparent distress HEENT: PERRL, EOMI, moist mucous membranes Cardiovascular: Regular rate, regular rhythm Pulmonary: Clear to auscultation bilaterally Abdomen: Soft, nontender, nondistended. No rebound tenderness. MSK: Moves extremities spontaneously Skin: No lesions on chest, face, upper extremities. Neuro: No gross focal neuro deficits Psych: Appropriate affect, thought content normal. Orientation Level: Oriented X7 Cognition: Follows commands Speech: Clear Vision: Functional Hearing: Functional Assistive Devices: None Data/Results Results are shown for the following tests if performed (CBC, Chem 7, Mg, Coag). If the patient did not have any of these tests, no results will be shown here. Lab Results Component Value Date/Time WBC 4.57 // 0545 RBC 5.35 // 0545 HGB 14.8 //0545 HCT 45.5 // 0545 MCH 27.7 //0545 MCV 85.0 //0545 PLT 157 //0545 RDW 12.7 // 0545 Lab Results Component Value Date/Time NA 138 //0545 K 3.6 // 0545 CL 98 // 0545 CO2 25 // 0545 BUN 7 // 0545 CRE 1.18 // 0545 CA 8.4 (L) // 0545 GLU 122 (H) // 0545

Table A4. Performance metrics used for model assessment.

Performance metrics

$$FPR = \frac{FP}{FP + TN} \; (A1)$$

$$Recall = \frac{TP}{TP + FN} \; (A2)$$

$$AP = \sum\nolimits_{k} (Recall_k - Recall_{k-1}) \, PPV_k \; (A3)$$

$$F1 = \frac{2TP}{2TP + FN + FP} \; (A4)$$

$$Accuracy: ACC = \frac{TN + TP}{TN + TP + FN + FP} \; (A5)$$

$$Precision = \frac{TP}{TP + FP} \; (A6)$$

TN and TP indicate the true negatives and true positives; FN and FP indicate the false negatives and false positives.

For the multiclass problem, for each class *c*:

$$TP_i = c_{ii}$$

$$FP_i = \sum\nolimits_{j=1}^{L} c_{ji} - TP_i$$

$$FN_i = \sum\nolimits_{j=1}^{L} c_{ij} - TP_i$$

$$TN_i = \sum\nolimits_{j=1}^{L} \sum\nolimits_{l=1}^{L} c_{jl} - TP_i - FP_i - FN_i$$

where each element i, j corresponds to the number of items with true class *i* that were classified as being in class j, with $j \in \{1,..,L\}$ and L as the total number of classes.

Appendix B

Figure B1. Model average training performance in 5-fold cross validation for different values of the inversed regularization strength constant C.

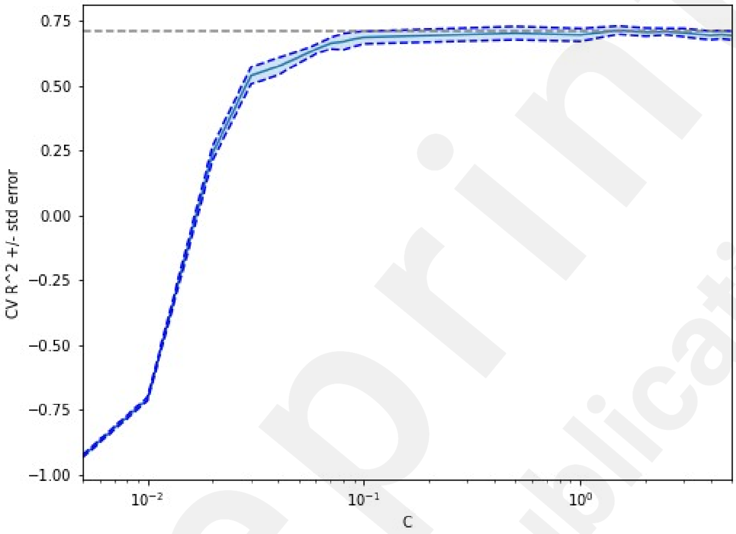
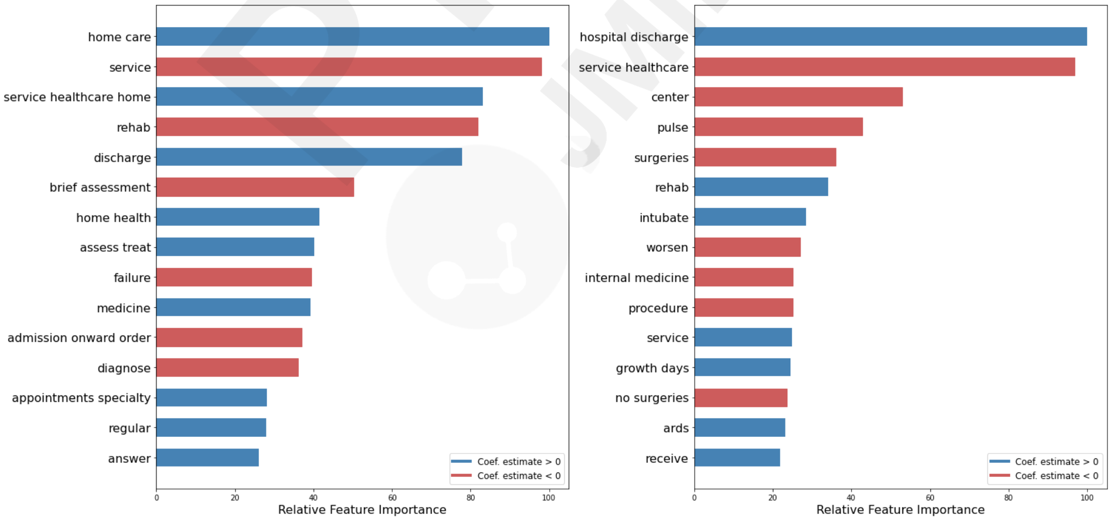
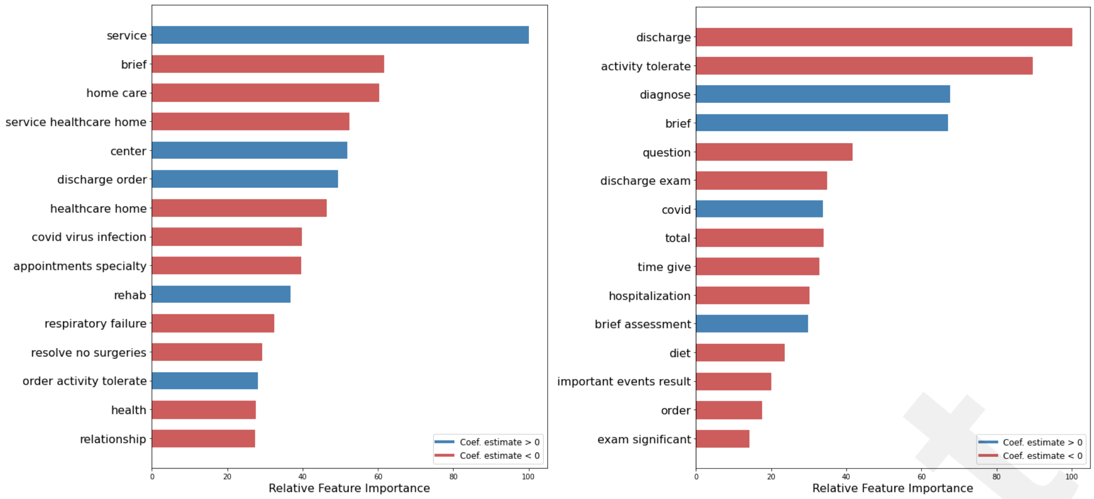


Figure B2. Relative importance of top 15 features obtained for the binary "one-vs-rest" model for each discharge outcome: (a) home; (b) inpatient rehabilitation; (c) skilled inpatient nursing facility; and (d) death.

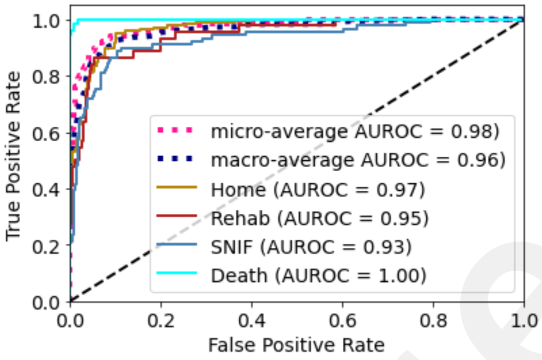


(a)(b)

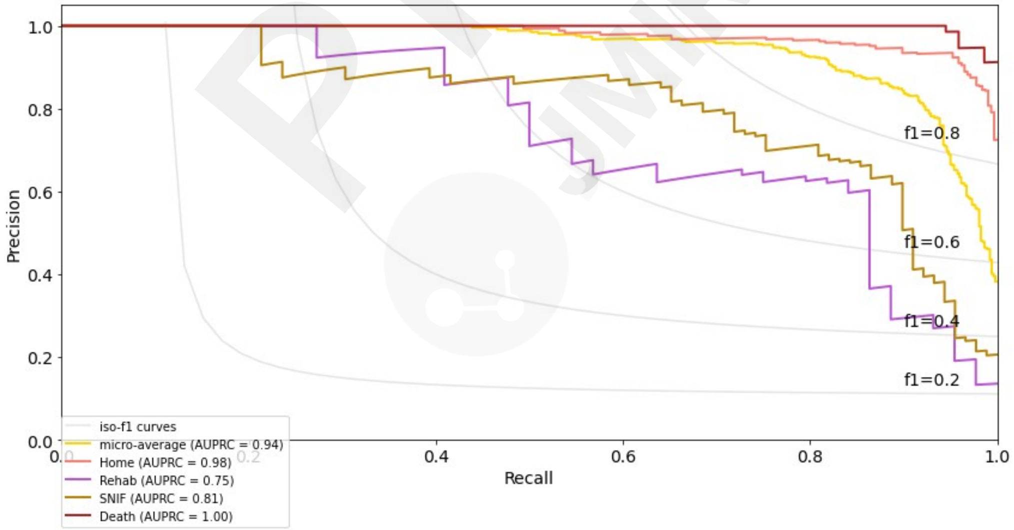


(c) (d)

Figure B3. Areas under the (a) ROC curve (AUROC), and (b) Precision-Recall curve (AUPRC), for the best model evaluated in the hold out test set.

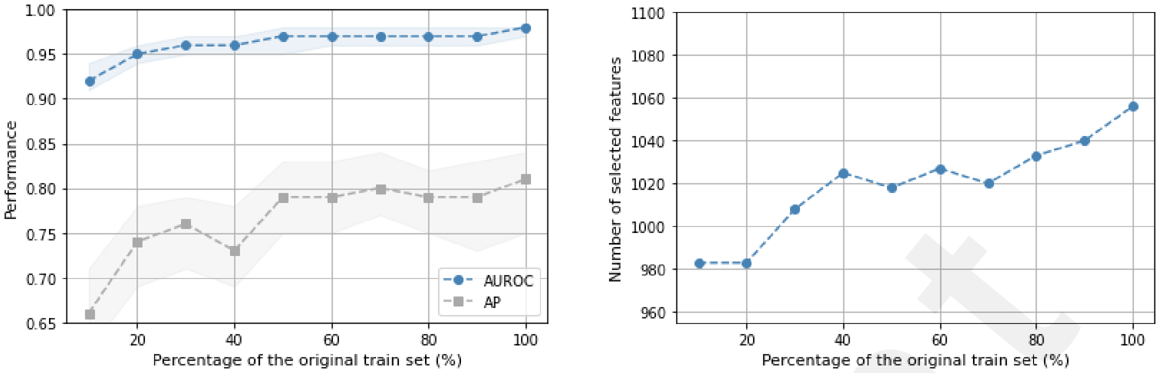


(a)



(b)

Figure B4. (a) Model performance for the best model evaluated in the hold out test set and (b) number of selected features in train, according to each train set (dimensions 10-100% of the original train set).



(a) (b)

Figure B5. Number of common features between the top 30 features selected for each train set (dimensions 10-100% of the original train set) and the top 30 features from the original train set.

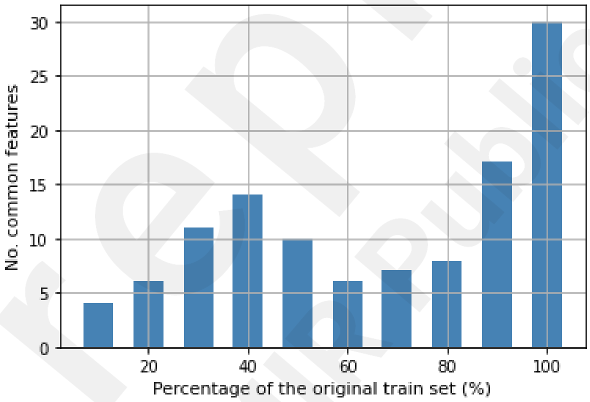
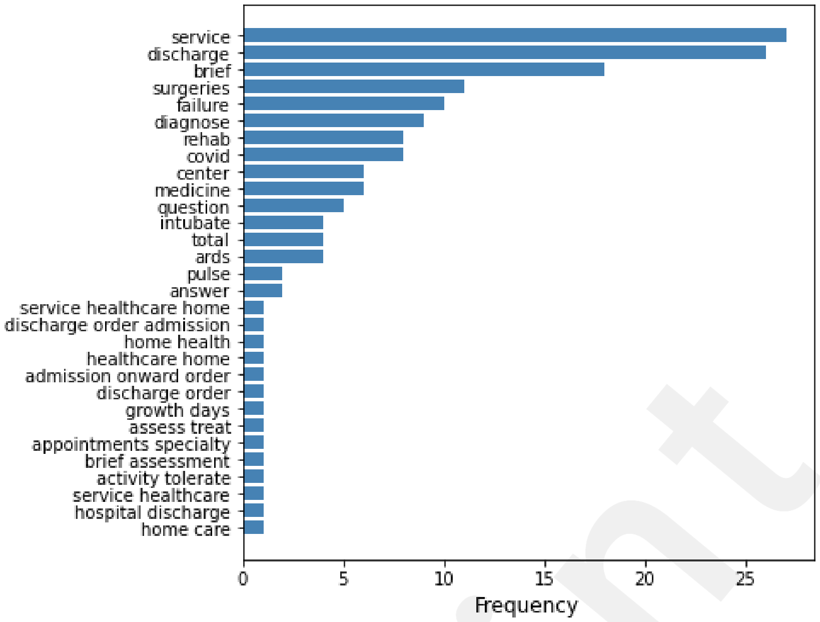


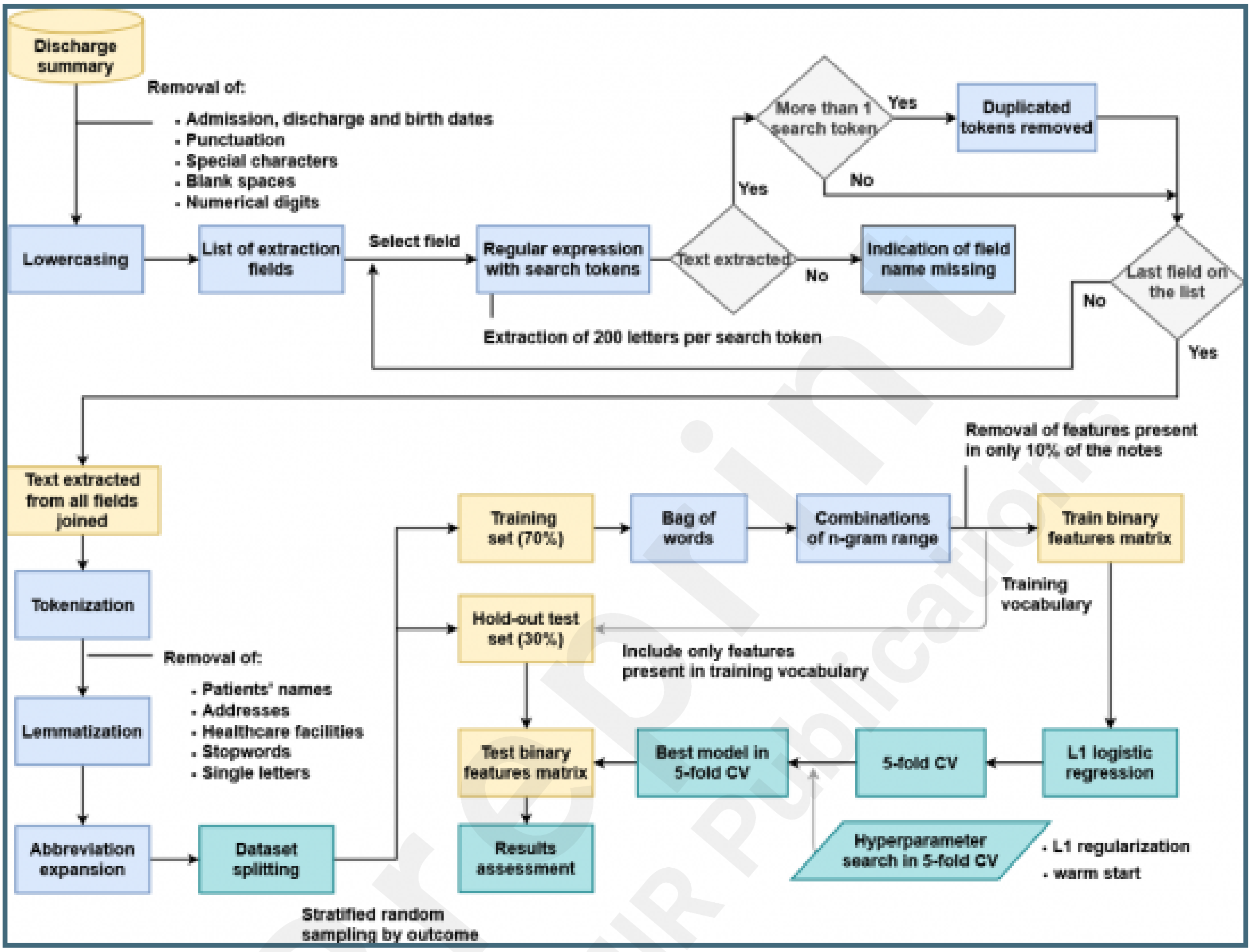
Figure B6. Frequency of features selected, among all train sets (dimensions 10-100% of the original train set), for the top 30 features selected from the original train set.


service
discharge
brief
surgeries
failure
diagnose
rehab
covid
center
medicine
question
intubate
total
ards
pulse
answer
service healthcare home
discharge order admission
home health
healthcare home
admission onward order
discharge order
growth days
assess treat
appointments specialty
brief assessment
activity tolerate
service healthcare
hospital discharge
home care
0
5
10
15
20
25
Frequency

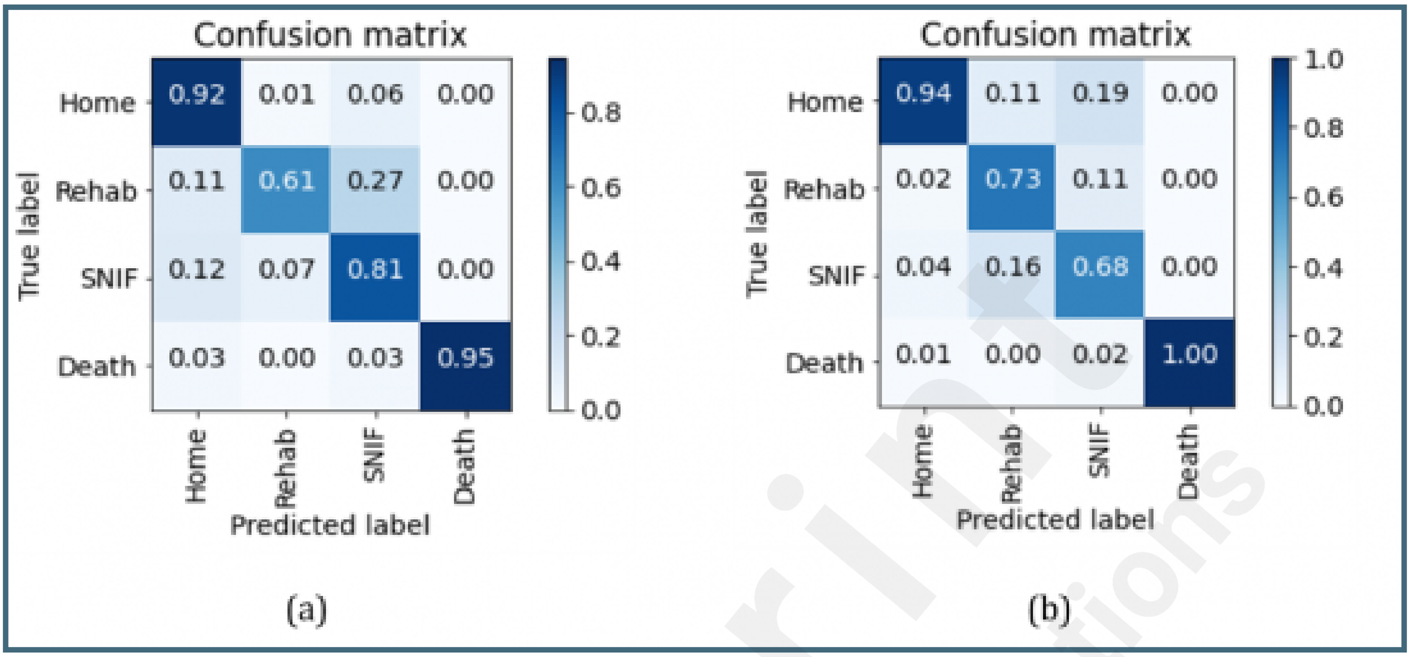
Supplementary Files

Figures

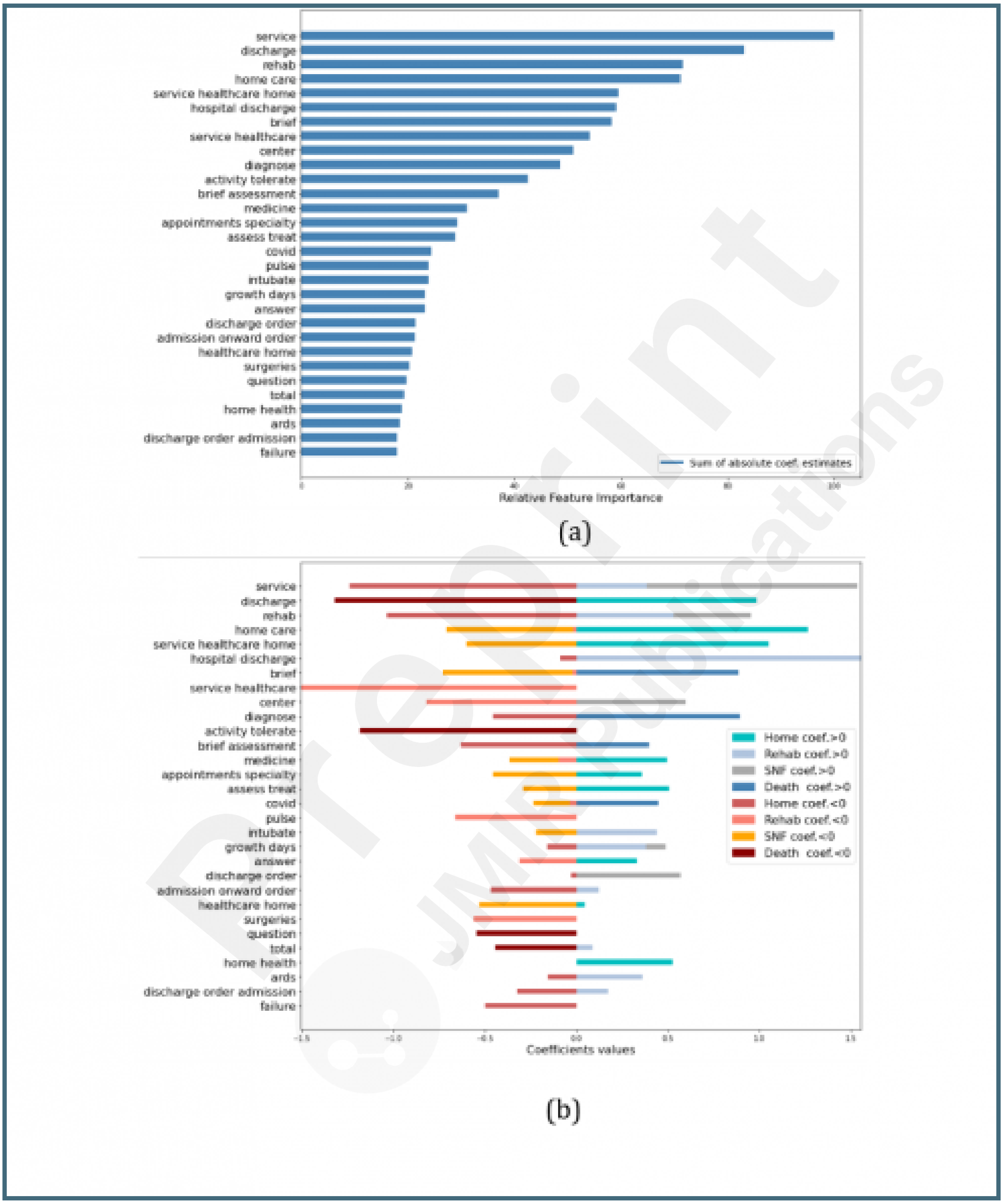
Methodology steps for discharge summary notes preprocessing and modeling. The list of extraction field is depicted in Table 1.



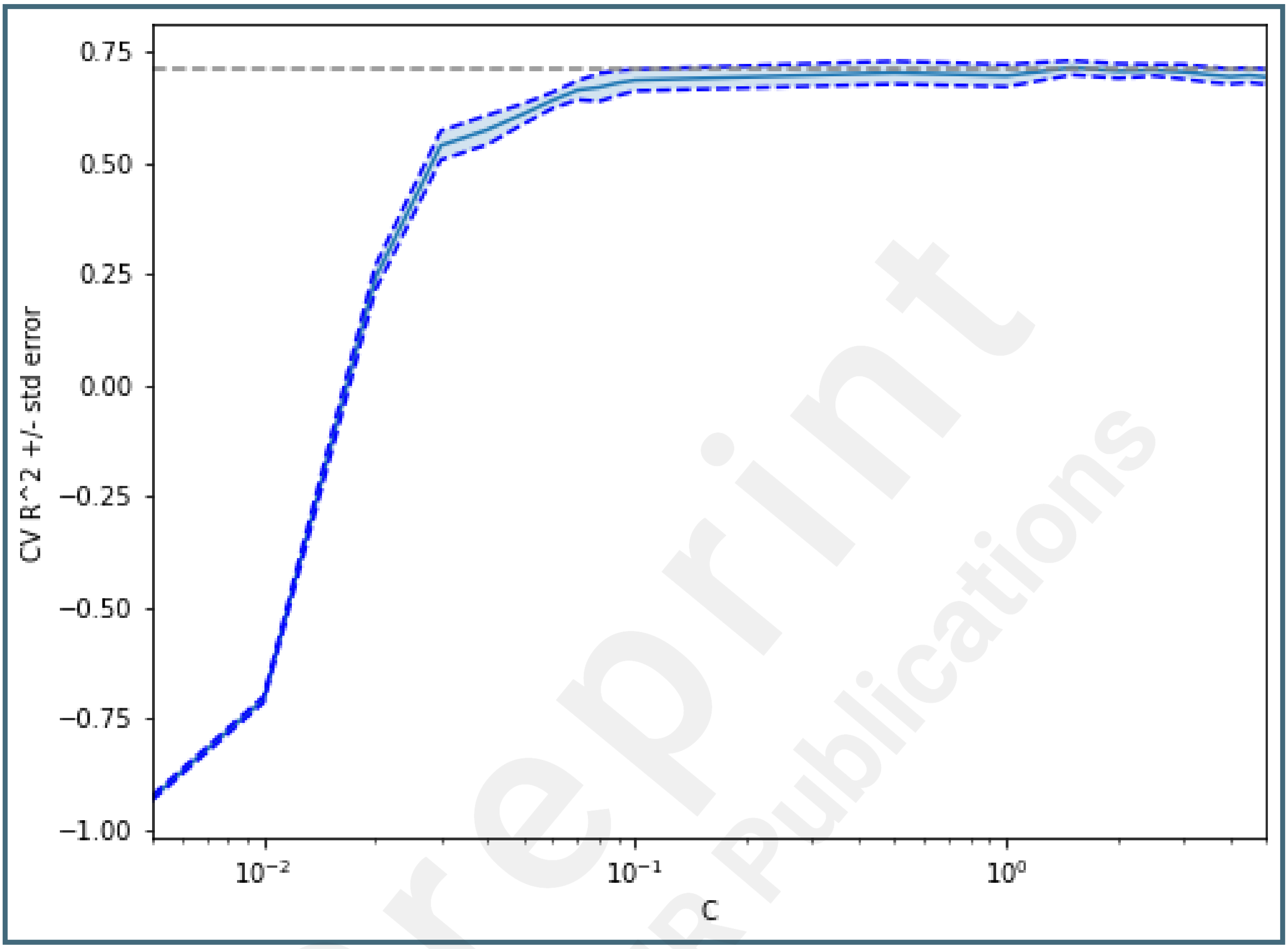
Confusion matrices for the best model evaluated in the hold-out test set normalized (a) by recall and (b) by precision.



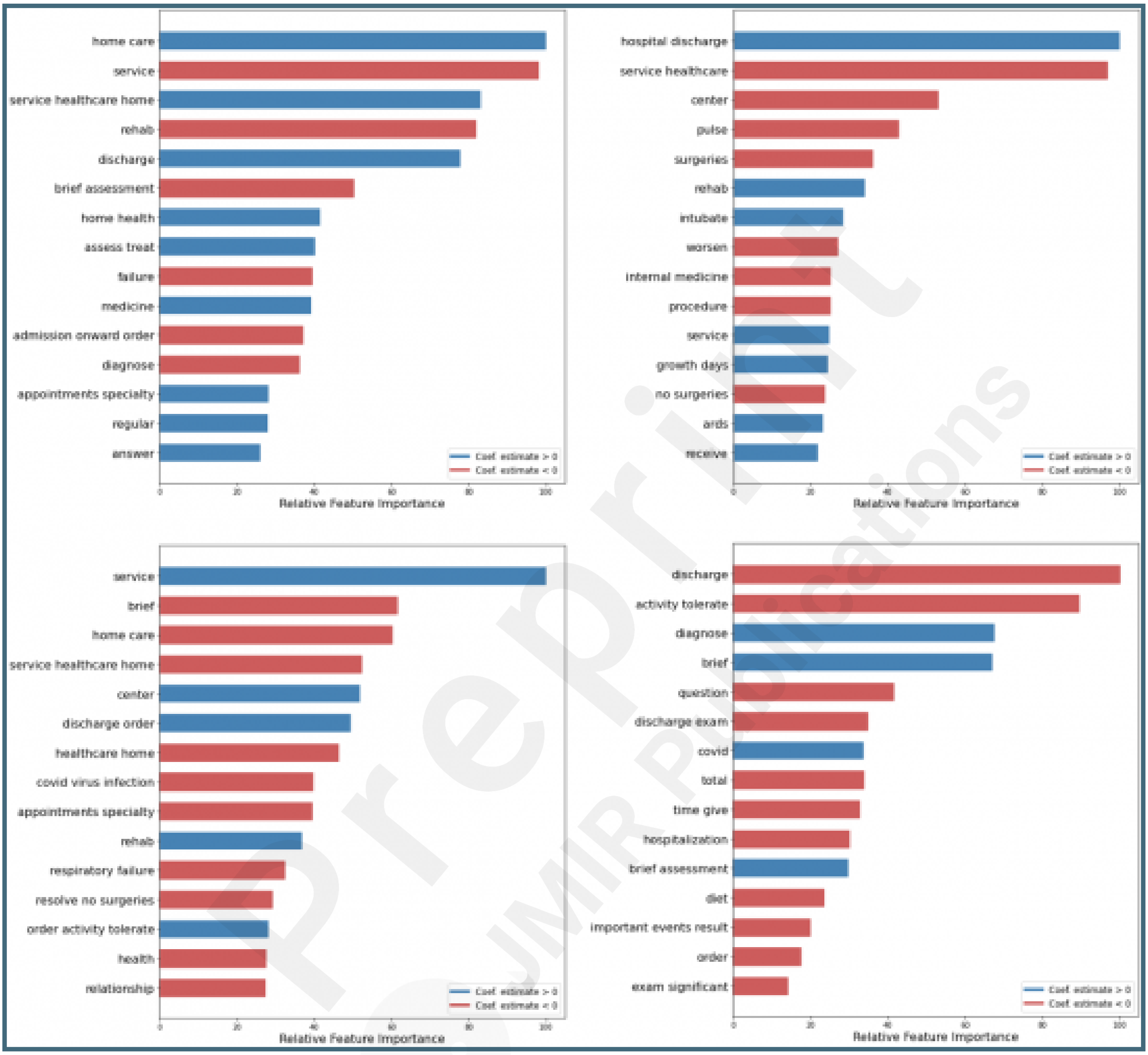
Relative importance of top 30 features obtained with the model coefficients estimates for (a) the sum of the absolute coefficients values and (b) the coefficients values discriminated by outcome.



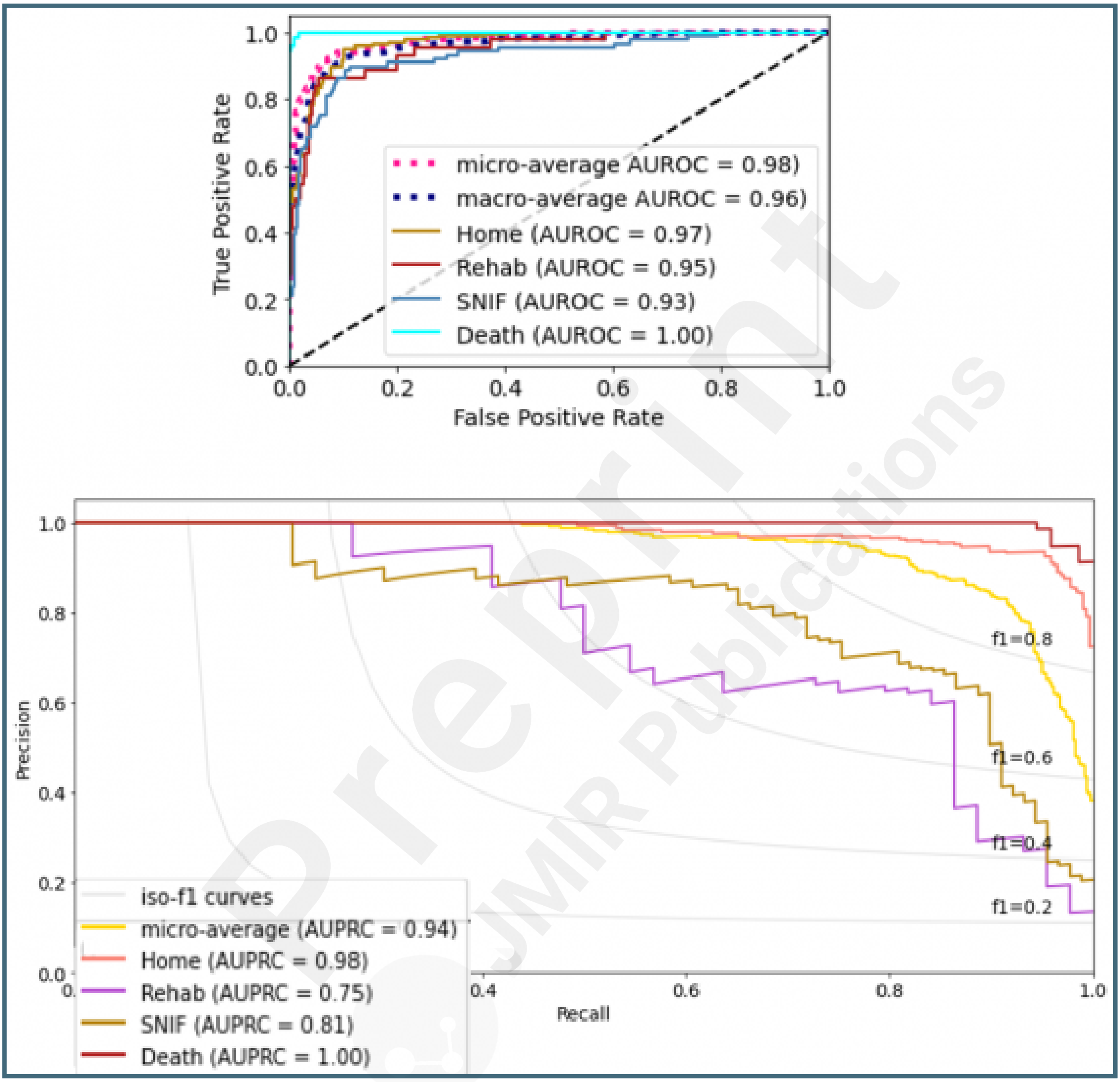
Model average training performance in 5-fold cross validation for different values of the inversed regularization strength constant C.



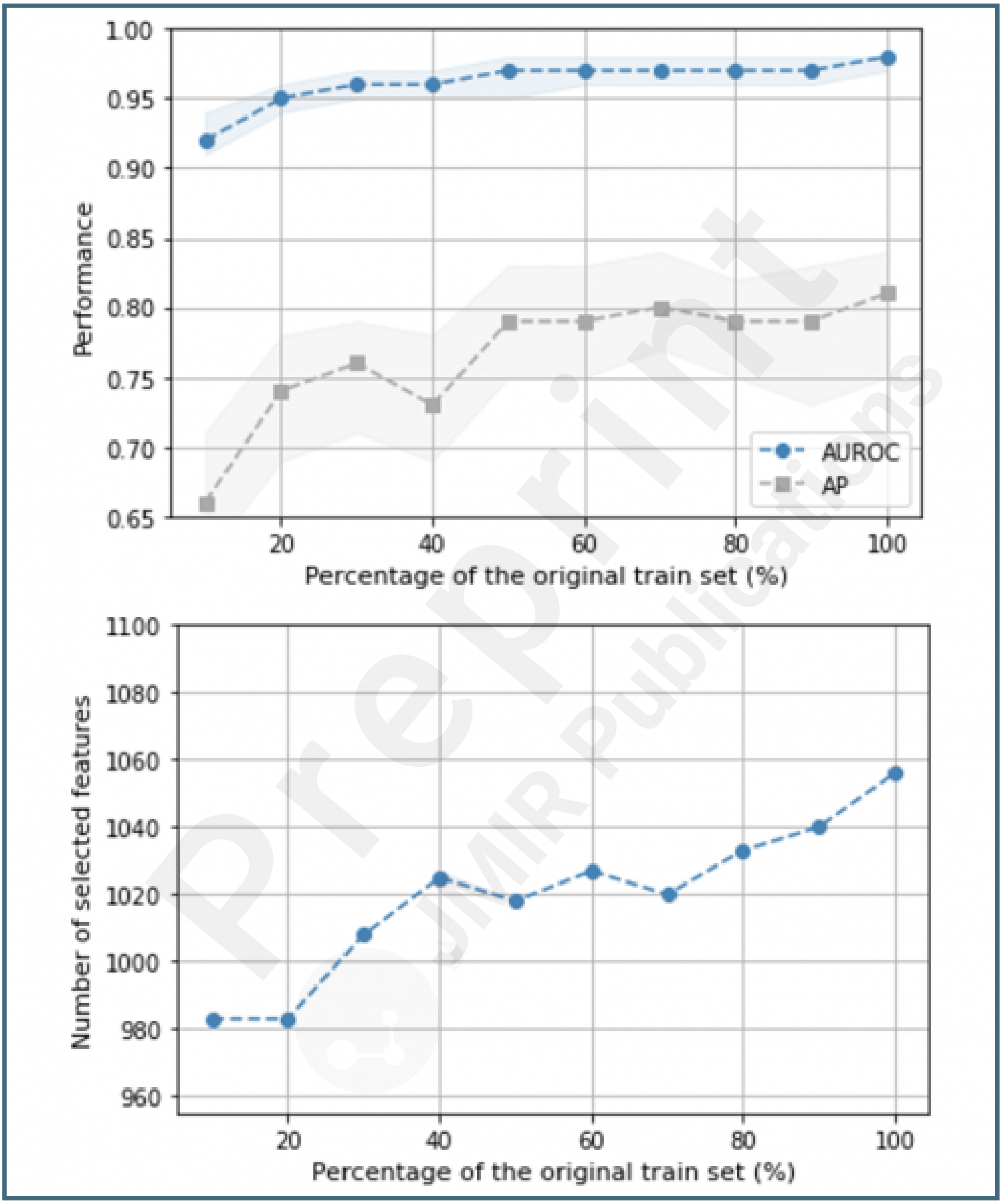
Relative importance of top 15 features obtained for the binary "one-vs-rest" model for each discharge outcome: (a) home; (b) inpatient rehabilitation; (c) skilled inpatient nursing facility; and (d) death.



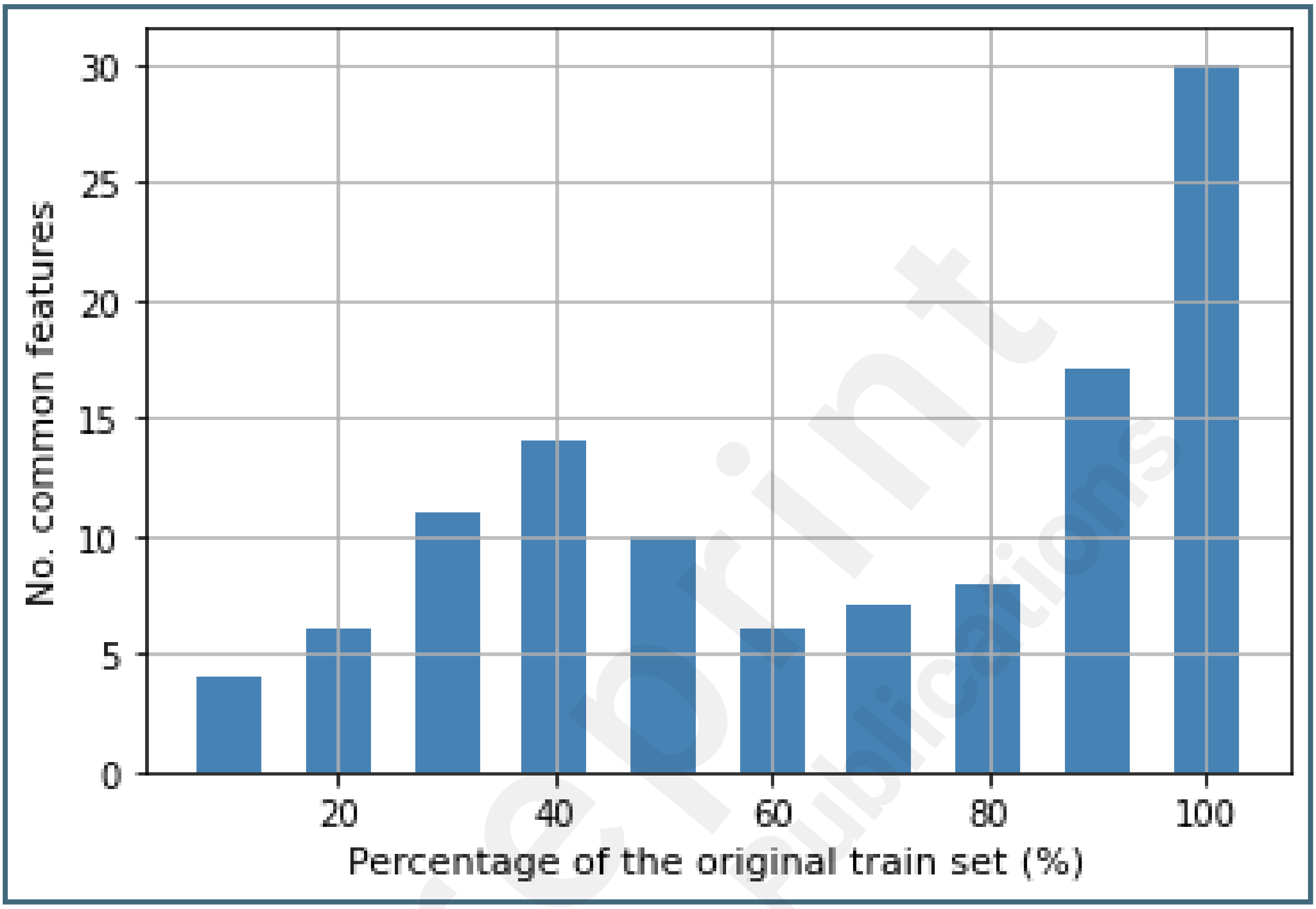
Areas under the (a) ROC curve (AUROC), and (b) Precision-Recall curve (AUPRC), for the best model evaluated in the hold out test set.



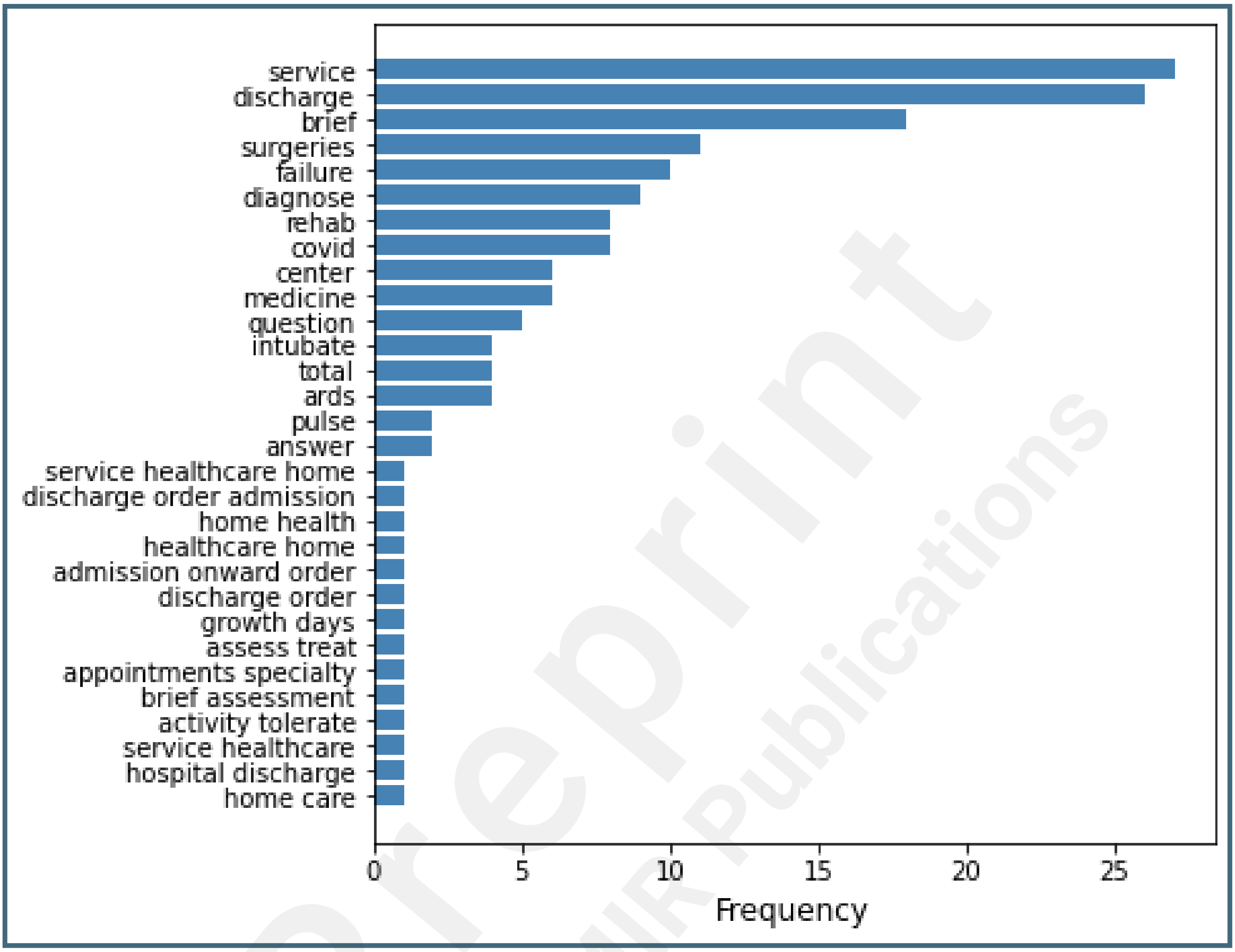
(a) Model performance for the best model evaluated in the hold out test set and (b) number of selected features in train, according to each train set (dimensions 10-100% of the original train set).



Number of common features between the top 30 features selected for each train set (dimensions 10-100% of the original train set) and the top 30 features from the original train set.



Frequency of features selected, among all train sets (dimensions 10-100% of the original train set), for the top 30 features selected from the original train set.



Multimedia Appendixes

Methodology.
URL: https://asset.jmir.pub/assets/915c9d733b18571e3328f28228b498e7.doc

Results.
URL: https://asset.jmir.pub/assets/302fdc06f06952967f197b20d7b566f5.doc