UC Davis
UC Davis Previously Published Works

Title

Predicting Mortality in Critical Care Patients with Fungemia Using Structured and Unstructured Data*

Permalink

ISBN

Authors

Baxter, Sally L
Klie, Adam R
Saseendrakumar, Bharanidharan Radha
et al.

Publication Date

DOI

Predicting Mortality in Critical Care Patients with Fungemia Using Structured and Unstructured Data

Sally L. Baxter, **Adam R. Klie**, **Bharanidharan Radha Saseendrakumar**, **Gordon Y. Ye**, **Michael Hogarth**, **Shamim Nemati [Member, IEEE]**

University of California San Diego (UCSD), La Jolla, CA 92093, USA

Abstract

Fungemia is a life-threatening infection, but predictive models of in-patient mortality in this infection are few. In this study, we developed models predicting all-cause in-hospital mortality among 265 fungemic patients in the Medical Information Mart for Intensive Care (MIMIC-III) database using both structured and unstructured data. Structured data models included multivariable logistic regression, extreme gradient boosting, and stacked ensemble models. Unstructured data models were developed using Amazon Comprehend Medical and BioWordVec embeddings in logistic regression, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). We evaluated models trained on all notes, notes from only the first three days of hospitalization, and models trained on only physician notes. The best-performing structured data model was a multivariable logistic regression model that achieved an accuracy of 0.74 and AUC of 0.76. Liver disease, acute renal failure, and intubation were some of the top features driving prediction in multiple models. CNNs using unstructured data achieved similar performance even when trained with notes from only the first three days of hospitalization. The best-performing unstructured data models used the Amazon Comprehend Medical document classifier and CNNs, achieving accuracy ranging from 0.99–1.00, and AUCs of 1.00. Therefore, unstructured data - particularly notes composed by physicians - offer added predictive value over models based on structured data alone.

Clinical Relevance—Unstructured data may help identify which patients are at high risk of death and need more aggressive treatment or closer monitoring. This is critically important since mortality rates from fungemia are very high.

I. Introduction

Fungal bloodstream infections (fungemia) rank among the top five hospital-acquired infections in many countries [1], [2]. Despites advances in anti-fungal therapies, fungemia is still associated with high mortality, which ranges from 35–75% [3]. Furthermore, it is associated with prolonged hospitalizations and high costs of care [4], [5]. Numerous models have been developed to predict which patients are at highest risk of developing fungemia [6]–[9]. These studies have identified risk factors such as immune suppression, recent major surgery, intravenous drug use, hyperalimentation, and indwelling central catheters. However, models predicting downstream end-organ damage and death among fungemic patients are

correspondence: s1baxter@health.ucsd.edu.

less common. Previously published models have utilized primarily multivariable logistic regression of structured data such as discrete clinical and laboratory variables [10]–[13]. However, despite significant advances in natural language processing (NLP) methods in recent years, to our knowledge no models predicting death in fungemic patients using unstructured data from free-text clinical notes have been published thus far.

To address this gap, we developed models predicting mortality among critical care patients with fungemia using both structured and unstructured data, hypothesizing that unstructured data would provide additional predictive value.

II. Methods

A. Study Population

The study population consisted of all patients with blood cultures positive for any fungal organisms in the most recent version of the Medical Information Mart for Intensive Care (MIMIC-III), a database of critical care patients at the Beth Israel Deaconess Medical Center (Boston, MA) [14]. This resulted in a cohort of 265 unique patients. MIMIC-III includes data from over 46,000 critical care patients (adults and neonates) from 2001–2012 [14]. Besides structured data such as demographics, laboratory values, diagnosis and procedure codes, microbiology data, medications, and physiologic measurements, it also includes de-identified unstructured data such as daily provider notes and imaging reports. The UCSD Institutional Review Board (IRB) determined that this project was non-human subjects research and would not require full IRB review. The research adhered to the tenets of the Declaration of Helsinki.

B. Data Curation and Processing

The outcome used in this study was all-cause in-hospital mortality during the hospitalization where positive fungal blood culture was documented. The following categories of structured data were extracted: demographics, comorbidities, procedures, lab values, fungal species identified in culture, and other risk factors known to be associated with fungemia. For unstructured data, free-text clinical notes were extracted. In addition to the text itself, we extracted the date and time of note creation as well as the category of note. We created one dataset containing all categories of notes (physician notes, nursing and other provider notes, imaging reports, etc.) and encompassed the entire hospitalization, from initial admission until discharge or death. This totaled 26,830 notes spanning the entire hospitalization. We also curated datasets containing subsets of the notes to evaluate differences in model performance based on timing and type of notes. Regarding note timing, we created the following note groups: (1) all notes from the first day of hospitalization (859 notes), (2) all notes from the first 2 days (i.e. notes from day 1 and day 2 combined, totaling 2137 notes), and (3) all notes from the first 3 days (3183 notes). Finally, we created a fifth dataset that included all notes labeled as physician notes in the entire hospitalization (1548 notes). For all models, training and testing were performed using a 90%/10% split of the data.

C. Predictive Modeling Using Structured Data

First, structured data were used as features for multivariable logistic regression models. An Automated Machine Learning (AutoML) framework (h2o) [15] was then used to perform to test alternative models. Hyperparameter tuning was performed through randomized grid search for Generalized Linear Models (GLMs), eXtreme Gradient Boosting (XG Boost), Gradient Boosting Machine (GBM), and deep learning models. Model performance ranking was determined through five-fold cross validation during the model tuning stage.

D. Predictive Modeling Using Unstructured Data

All five unstructured datasets (all notes, 3 subsets of notes based on timing, and all physician notes) were used to train the various models described below.

Amazon Comprehend Medical: Amazon Comprehend Medical is an NLP service offered by Amazon Web Services (AWS) that leverages machine learning techniques to extract relevant medical information from unstructured text [16]. The service uses a pre-trained model offered as a document classifier to examine and analyze unstructured data. We used Amazon Comprehend Medical's document classifier to create a binary classification model with in-hospital mortality as the label and clinical notes as input documents. Before feeding the documents to the classifier, punctuations were removed. The clinical notes were grouped by patient and concatenated to build a single document per patient, except for physician notes which remained ungrouped.

Models Utilizing BioWordVec Embeddings: In order to build mortality classification models that take the five aforementioned unstructured datasets as input, numerical vector representations of clinical notes were obtained using a pre-trained model. Clinical notes extracted from MIMIC-III were first tokenized using the NLTK (v3.4.4) Python package [17], and stop words and punctuation were removed. Tokens for each note were vectorized into 200-dimensional vectors using the BioWordVec subword embedding model from Zhang *et al.* [18]. The BioWordVec model utilizes a skip-gram approach that represents each word as a sum of its character n-grams [19]. The BioWordVec model was pre-trained on over 28,000,000 PubMed documents and over 2,000,000 clinical notes from MIMIC-III. As input to a logistic regression model, token vectors for a given note were averaged to produce a mean 200-dimensional embedding vector. These mean embedding vectors were used as input to train logistic regression models from python's Scikit-learn (v0.22) package [20]. CNNs and RNNs were trained on each dataset using a max length of 400 and 100 respectively. The neural nets were implemented with Python's Keras (v2.3.1) API [21]. All CNNs included a 1D convolutional layer with 16 filters, a global max pooling layer, a dense 250-dimensional hidden layer and a dropout rate of 0.5. The RNNs included a simple recurrent layer with 50 neurons and a dropout rate of 0.2. All neural networks were trained using 10 epochs and a batch size of 32.

E. Performance Metrics

For all models, the following were used to evaluate performance: area under the receiver operating characteristic curve (AUC), accuracy, precision, recall, and F1 score. Receiver operating characteristic curves could not be generated using the Amazon Comprehend

Medical platform, so confusion matrices were generated instead to evaluate performance. Of note, the metrics provided by Amazon Comprehend Medical are macro averages, i.e., accuracy, precision, recall and F1-score are averaged over both labels rather than on only the positive label.

F. Masking Assessment of Unstructured Data Models

To identify which concepts may be driving predictive performance in the unstructured data models, particularly for physician notes, masked datasets were created. Each masked dataset had terms removed related to the following concepts hypothesized to influence prediction: central catheterization, diabetes, cancer, chemotherapy, steroids, intravenous drug use, hyperalimentation, acute renal failure, hemodialysis, liver disease, and intubation. For example, the dataset masking the significant co-morbidity of diabetes had the following terms removed: "diabetes", "dm", "dm1", "dm2", "t1dm", "t2dm", "diabetic", "dka", "insulin", "metformin", "glipizide", "glyburide", "lantus", and "humalog." The masked datasets were then each used to train models using the Amazon Comprehend Medical classifier, logistic regression, CNNs, and RNNs. Concepts masked from datasets resulting in models with reduced performance were inferred to be important for prediction.

In addition, to ensure that predictions were not simply based on physician notes expressing imminent death of the patient (particularly at the end of the hospitalization immediately preceding death), a final masked dataset was created where terms related to death were removed. These included "dying", "death", "die", "mortality", "dnr", "dni", and "palliative".

III. Results

A. General Features of the Study Population

A total of 265 patients were identified from the MIMIC-III database as having fungemia, with documented blood cultures positive for fungal organisms. The mean (SD) age was 62.3 (16.8) years. There was a slight male predominance (154/265, 58%), and the cohort was majority white (199/265, 75%). The mean (SD) length of stay was 24.6 (27) days. The most common fungal species found on blood cultures were *Candida albicans* (43%), *Candida glabrata* (28%), *Candida parapsilosis* (12%) and *Candida tropicalis* (8%). Of the 265 patients, 121 died during the hospitalization, constituting an in-hospital mortality rate of 46%. Table 1 describes selected features among fungemic patients included in the study.

B. Performance of Structured Data Models

Multivariable Logistic Regression Models: The multivariable logistic regression (LR) model included structured data features chosen based on existing literature. This model had an AUC of 0.76 and accuracy of 0.74 (Table 2). The top features driving prediction were liver disease, acute renal failure, intubation, treatment with amphotericin B, cancer, infection with *Candida parapsilosis*, indwelling central catheter placement, recent major abdominal surgery, and treatment with micafungin.

AutoML Framework Models: Out of the various models trained using the AutoML framework, the two best-performing models were a Stacked Ensemble of binomial GLMs

and an XGBoost model. Both the Stacked Ensemble model and the XGBoost model demonstrated similar AUCs, at 0.67. The Stacked Ensemble demonstrated superior accuracy at 0.74, compared with XGBoost at 0.67 (Table 2). The top features driving prediction in the XGBoost model were age, renal function, acute renal failure, intubation, liver disease, female gender, infection with *Candida albicans*, diabetes, infection with *Candida glabrata*, and fluconazole. Due to technical limitations of the h2o AutoML framework, the best performing model (Stacked Ensemble) was not interpretable, and factors driving prediction for that specific model could not be determined.

C. Performance of Unstructured Data Models

Unstructured Datasets: The dataset containing all notes from the entire hospitalization of patients in our cohort consisted of 26,830 notes. Unsurprisingly, the dataset limited to the first day of hospitalization was the smallest (859 notes), and as subsequent datasets included more hospital days, the number of notes also increased. There were 1548 notes categorized as physician notes. Notes were evenly balanced between patients who died and those who survived in all datasets.

Results of Amazon Comprehend Medical Classifier: Amazon Comprehend Medical's classifier performed poorly on the grouped clinical notes from day 1 of hospitalization (Table 3). Although the accuracy on test data is within reasonable limits compared to the dataset with notes on days 1 and 2 of hospitalization, the model reports no true positives (0% TPR - True Positive Rate). The model built on notes from day 1 and 2 of hospitalization reports more false positives than true positives from test data (27% TPR). With notes from the first 3 days of hospitalization, the reports higher percentage of true positives on the test data compared to the previous dataset (36% TPR). In general, TPR increases as more days of hospitalization (and hence greater document size) are included. However, the model built on physician notes, although not the biggest dataset, outperforms all the other datasets, reporting 100% on all metrics and 100% TPR.

Results of Models using BioWordVec Embeddings: Word embeddings generated from the unstructured clinical notes and the BioWordVec model were used to train and test a logistic regression, a CNN and an RNN for each dataset (Table 3, Figs. 1 and 2). CNNs achieved the highest F1 and AUC across all five datasets. The CNN trained on notes from only the first 3 days of hospitalization achieved comparable performance metrics (Table 3) to the best structured data model (Table 2) and much better performance than the Amazon Comprehend Medical document classifier on the same dataset (Table 3). Including all notes available for fungemic patients resulted in boosted performance for all three model types, with the CNN showing significantly better performance than the best structured data model and comparable performance to the Amazon Comprehend Medical document classifier. Furthermore, the overall trend of increasing performance with increasing data size observed with the Amazon Comprehend Medical document classifier was recapitulated for BioWordVec based models. Performance on notes from the first 2 days of hospitalizations was relatively poor, with the CNN trained on notes from within 2 days of hospitalization being the only model with classification performance comparable to the structured data models. These results suggest the potential of using clinical notes and sequential predictive

models, especially CNNs, for predicting mortality among fungemic patients, even when limiting notes to the first 2–3 days of hospitalization. As seen with the Amazon Comprehend Medical document classifier, physician notes resulted in almost perfect classification accuracy and AUC in both the CNN and RNN (Fig. 1), and a very high accuracy of 0.87 when using logistic regression (Table 3). These results indicate that physician notes may offer better predictive power than other note types available in MIMIC-III.

D. Masking Assessment of Unstructured Data Models

The performance of Amazon Comprehend Medical's document classifier remained excellent across all masked models (accuracy of 1.00 and F1 of 1.00), despite removal of terms related to concepts describing risk factors or death. CNNs still performed impressively, with accuracy ranging from 0.97–0.99 and F1 ranging from 0.98–0.99 even with masking. Only RNNs showed some decreased performance with masking, with accuracy decreasing to 0.91 when the terms related to intubation, hemodialysis, and cancer were removed. Even when terms directly related to death were removed from the physician notes, performance of all models were comparable to those using the original unmasked physician notes as inputs (accuracy of 0.84 for logistic regression, 0.98 for CNNs, and 0.95 for RNNs).

IV. Discussion and Future Directions

Here we describe multiple models predicting mortality in critical care patients with fungemia, using both structured and unstructured data. The overall best-performing structured data model was a multivariable logistic regression model achieving an accuracy of 0.74 and an AUC of 0.76. Out of prior models examining mortality in fungemic patients [10–13], only one study [11] reported AUC, which was 0.74 in a historical dataset. We used static data rather than physiologic time-series data for ease of comparison to existing models that utilized static data. Incorporation of structured time-series data represents an interesting area of future study. However, structured data still comprise only a small fraction of information regarding a patient's hospitalization, so we also trained several models using unstructured data and achieved superior predictive performance. To our knowledge, this is the first time natural language processing of unstructured notes has been used alongside structured data in a machine-learning based mortality prediction model for patients with fungemia.

The Amazon Comprehend Medical classifier was an easy-to-use, "off-the-shelf" clinical NLP system that performed well with all categories of notes from the hospitalization and also achieved high performance across all metrics when using all physician notes. A key advantage was rapid implementation without the need for custom programming. Another advantage of the service was the availability of a scalable environment to train models. The model trained with the largest dataset (all 26,830 notes from the hospitalization for patients in the cohort) took a run time (train + test) of 45 minutes. However, one of the system's current limitations is the lack of direct access to the classifier's model. As a result, we were unable to evaluate the parameters used to develop the model. Although Amazon did provide performance metrics and confusion matrices, there was not enough information provided to

generate receiver operating characteristic curves to allow comparisons of AUCs against other models.

In contrast, we manually developed logistic regression models, CNNs, and RNNs using BioWordVec embedding. Though the design and implementation of these approaches were not as straightforward as Amazon Comprehend Medical, open-source programming packages such as *keras* and *sci-kit learn* allowed for a relatively straightforward model building and testing pipeline. Furthermore, training and testing the sequential neural network models took <30 minutes on a 1.6 GHz Intel Core i5 processor for RNNs and CNNs on the largest dataset and <5 minutes for the CNN on physician notes. Logistic regression models trained with unstructured data performed poorly, whereas CNNs generally performed the best. Even when only including unstructured data from the first three days of hospitalization, the CNN performed reasonably well (accuracy of 0.68, AUC of 0.75) at predicting whether a fungemic patient would ultimately die during the admission. This was comparable to the best-performing structured data models, which utilized data from the entire hospitalization. Looking at data from the first three days of hospitalization, the CNN also substantially outperformed the Amazon Comprehend Medical document classifier. This is interesting given that fungal cultures typically take days to become positive, so the unstructured notes in the first 3 days may not even mention fungemia or anti-fungal therapeutics. The predictive power of CNNs using data from only the first three days of hospitalization is also significant from the standpoint of clinical decision support. For clinical decision support tools aimed at identifying which fungemic patients are at highest risk of death, it would not be useful to provide a prediction at the end of the hospitalization. Rather, estimation of risk early in the hospitalization is far more helpful.

Physician notes were associated with the best predictive performance across all model types. We performed several follow-up analyses to ensure the validity of this finding. First, we determined that these results were not simply due to class imbalance, as the outcome labels for the physician notes dataset was divided equally between those who died (50%) and those who survived (50%). Additionally, to counter the fact that some physician notes might explicitly mention an impending death, we trained models using a dataset where terms describing or directly related to death were removed. Even with these terms removed, models trained with physician notes still achieved high performance. The performance of models based on RNNs decreased when terms related to intubation, hemodialysis, and cancer were removed, implying that these entities were relatively more important in driving the prediction of death compared to other entities. Interestingly, for the Amazon Comprehend Medical document classifier and CNNs, performance remained robust despite masking a wide array of terms. Currently, structured data analyses still offer superior interpretability, but future studies to enhance interpretability of the unstructured data models may improve the understanding of which unstructured features are driving prediction and provide additional clinical insights.

V. Conclusion

Predictive models using unstructured, free-text clinical notes provide superior performance for predicting mortality among critical care patients with fungemia compared with models

trained with only structured data. Applying natural language processing methods to physician notes offers a high-performing and efficient approach for mortality prediction in fungemia. Ultimately, using these models, clinical decision support tools could be developed to reliably predict which patients with fungemia are at risk of dying during the hospitalization. These tools could help better inform their management, enabling a precision medicine approach to treating patients with fungemia.
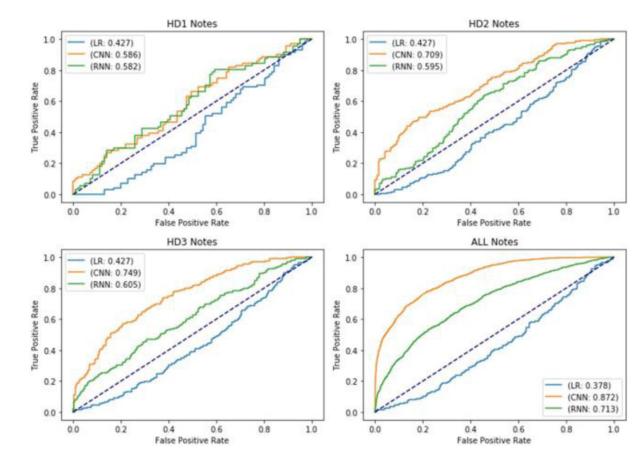
Acknowledgment

References

[1]. Magill SS et al., "Multistate point-prevalence survey of health care-associated infections," N. Engl. J. Med, vol. 370, no. 13, pp. 1198–1208, Mar. 2014, doi: 10.1056/NEJMoa1306801. [PubMed: 24670166]

[2]. Vincent J-L et al., "International study of the prevalence and outcomes of infection in intensive care units," JAMA, vol. 302, no. 21, pp. 2323–2329, Dec. 2009, doi: 10.1001/jama.2009.1754. [PubMed: 19952319]

[3]. Paiva J-A et al., "Characteristics and risk factors for 28-day mortality of hospital acquired fungemias in ICUs: data from the EUROBACT study," Crit Care, vol. 20, 2016, doi: 10.1186/s13054-016-1229-1.

[4]. Rajendran R et al., "A Prospective Surveillance Study of Candidaemia: Epidemiology, Risk Factors, Antifungal Treatment and Outcome in Hospitalized Patients," Front. Microbiol, vol. 7, 2016, doi: 10.3389/fmicb.2016.00915.

[5]. Wey SB, Mori M, Pfaller MA, Woolson RF, and Wenzel RP, "Hospital-acquired candidemia. The attributable mortality and excess length of stay," Arch. Intern. Med, vol. 148, no. 12, pp. 2642–2645, Dec. 1988, doi: 10.1001/archinte.148.12.2642. [PubMed: 3196127]

[6]. Fraser VJ, Jones M, Dunkel J, Storfer S, Medoff G, and Dunagan WC, "Candidemia in a tertiary care hospital: epidemiology, risk factors, and predictors of mortality," Clin. Infect. Dis, vol. 15, no. 3, pp. 414–421, Sep. 1992, doi: 10.1093/clind/15.3.414. [PubMed: 1520786]

[7]. Wenzel RP, "Nosocomial candidemia: risk factors and attributable mortality," Clin. Infect. Dis, vol. 20, no. 6, pp. 1531–1534, Jun. 1995, doi: 10.1093/clinids/20.6.1531. [PubMed: 7548504]

[8]. Gaspar GG, Menegueti MG, Auxiliadora-Martins M, Basile-Filho A, and Martinez R, "Evaluation of the predictive indices for candidemia in an adult intensive care unit," Rev. Soc. Bras. Med. Trop, vol. 48, no. 1, pp. 77–82, Feb. 2015, doi: 10.1590/0037-8682-0292-2014. [PubMed: 25860468]

[9]. Atamna A et al., "Predicting candidemia in the internal medicine wards: a comparison with gram-negative bacteremia-a retrospectives study," Diagn. Microbiol. Infect. Dis, vol. 95, no. 1, pp. 80–83, Sep. 2019, doi: 10.1016/j.diagmicrobio.2019.04.007. [PubMed: 31129007]

[10]. Vaquero-Herrero MP et al., "The Pitt Bacteremia Score, Charlson Comorbidity Index and Chronic Disease Score are useful tools for the prediction of mortality in patients with Candida bloodstream infection," Mycoses, vol. 60, no. 10, pp. 676–685, Oct. 2017, doi: 10.1111/myc.12644. [PubMed: 28833577]

[11]. Keighley C et al., "Candidaemia and a risk predictive model for overall mortality: a prospective multicentre study," BMC Infect. Dis, vol. 19, no. 1, p. 445, May 2019, doi: 10.1186/s12879-019-4065-5. [PubMed: 31113382]

[12]. Wang T-Y et al., "The clinical outcomes and predictive factors for in-hospital mortality in non-neutropenic patients with candidemia," Medicine (Baltimore), vol. 95, no. 23, p. e3834, Jun. 2016, doi: 10.1097/MD.0000000000003834. [PubMed: 27281087]
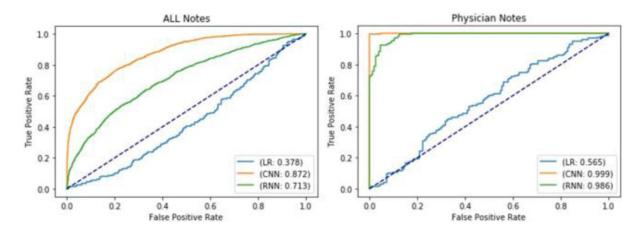
[13]. Jia X, Li C, Cao J, Wu X, and Zhang L, "Clinical characteristics and predictors of mortality in patients with candidemia: a six-year retrospective study," Eur. J. Clin. Microbiol. Infect. Dis, vol. 37, no. 9, pp. 1717–1724, Sep. 2018, doi: 10.1007/s10096-018-3304-9. [PubMed: 30030692]

[14]. Johnson AEW et al., "MIMIC-III, a freely accessible critical care database," Sci Data, vol. 3, no. 1, p. 160035, Dec. 2016, doi: 10.1038/sdata.2016.35. [PubMed: 27219127]

[15]. h2oai/h2o-3. H2O.ai, 2019.

[16]. "Amazon Comprehend Medical," Amazon Web Services, Inc. [Online]. Available: https://aws.amazon.com/comprehend/medical/. [Accessed: 11-Dec-2019].

[17]. Bird S, Klein E, and Loper E, Natural language processing with Python, 1st ed Beijing; Cambridge [Mass.]: O'Reilly, 2009.

[18]. Zhang Y, Chen Q, Yang Z, Lin H, and Lu Z, "BioWordVec, improving biomedical word embeddings with subword information and MeSH," Sci Data, vol. 6, no. 1, p. 52, Dec. 2019, doi: 10.1038/s41597-019-0055-0. [PubMed: 31076572]

[19]. Bojanowski P, Grave E, Joulin A, and Mikolov T, "Enriching Word Vectors with Subword Information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, Dec. 2017, doi: 10.1162/tacl_a_00051.

[20]. Pedregosa F et al., "Scikit-learn: Machine Learning in Python," MACHINE LEARNING IN PYTHON, p. 6.

[21]. keras-team/keras. Keras, 2019.

Fig 1.
Receiver operating characteristic curves for models predicting mortality trained on unstructured data of fungemic patients from the first day of hospitalization (HD1), the first 2 days of hospitalization (HD2), the first three days of hospitalization (HD3), and all notes (ALL).

Fig 2.
Receiver operating characteristic curves for models predicting mortality trained on all notes
(left) and physician notes only (right).

TABLE I.

Patients with Fungemia in MIMIC-III.

	Survived (n=144)	Died (n=121)	Overall (n=265)
GENDER			
Female	57 (39.6%)	54 (44.6%)	111 (41.9%)
Male	87 (60.4%)	67 (55.4%)	154 (58.1%)
ETHNICITY			
White	111 (77.1%)	88 (72.7%)	199 (75.1%)
Asian	4 (2.8%)	2 (1.7%)	6 (2.3%)
Black/African American	12 (8.3%)	10 (8.3%)	22 (8.3%)
Hispanic or Latino	3 (2.1%)	2 (1.7%)	5 (1.9%)
Other	2 (1.4%)	1 (0.8%)	3 (1.1%)
Patient declined or unknown	12 (8.3%)	18 (14.9%)	30 (11.3%)
MEAN (SD) AGE	60.1 (17.3)	65.0 (15.9)	62.3 (16.8)
CLINICAL FEATURES			
Indwelling Central Catheter	45 (31.2%)	51 (42.1%)	96 (36.2%)
Diabetes	35 (24.3%)	34 (28.1%)	69 (26.0%)
Cancer	8 (5.6%)	11 (9.1%)	19 (7.2%)
Chemotherapy	3 (2.1%)	1 (0.8%)	4 (1.5%)
Steroids	4 (2.8%)	3 (2.5%)	7 (2.6%)
Intravenous drug use	8 (5.6%)	3 (2.5%)	11 (4.2%)
Hyperalimentation	1 (0.7%)	1 (0.8%)	2 (0.8%)
Acute renal failure	67 (46.5%)	87 (71.9%)	154 (58.1%)
Hemodialysis	2 (1.4%)	2 (1.7%)	4 (1.5%)
Liver Disease	14 (9.7%)	29 (24.0%)	43 (16.2%)
Intubation	90 (62.5%)	94 (77.7%)	184 (69.4%)

TABLE II.

Performance of structured data models.

Model	Accuracy	Precision	Recall	F1	AUC
Multivariable Logistic Regression	0.74	0.73	0.67	0.70	0.76
Stacked Ensemble	0.74	0.83	0.67	0.36	0.67
XG Boost	0.67	0.92	0.58	0.36	0.67

TABLE III.

Performance of unstructured data models.

Model	Accuracy	Precision	Recall	F1	AUC
Amazon Comprehend Medical					
Hospital Day 1	0.57	0.28	0.50	0.36	-
Hospital Days 1–2	0.56	0.54	0.53	0.5	-
Hospital Days 1–3	0.48	0.47	0.47	0.46	-
All Notes	0.80	0.80	0.80	0.79	-
Physician Notes	1.00	1.00	1.00	1.00	-
Logistic Regression					
Hospital Day 1	0.58	0.48	0.35	0.41	0.43
Hospital Days 1–2	0.60	0.61	0.48	0.54	0.43
Hospital Days 1–3	0.57	0.55	0.52	0.54	0.43
All Notes	0.70	0.65	0.52	0.58	0.38
Physician Notes	0.87	0.85	0.91	0.88	0.57
Convolutional Neural Network (CNN)					
Hospital Day 1	0.53	0.44	0.48	0.46	0.59
Hospital Days 1–2	0.63	0.63	0.60	0.61	0.71
Hospital Days 1–3	0.68	0.66	0.65	0.66	0.75
All Notes	0.79	0.74	0.72	0.73	0.87
Physician Notes	0.99	1.00	1.00	0.99	1.00
Recurrent Neural Network (RNN)					
Hospital Day 1	0.55	0.46	0.44	0.45	0.58
Hospital Days 1–2	0.57	0.56	0.53	0.55	0.60
Hospital Days 1–3	0.57	0.55	0.51	0.53	0.61
All Notes	0.68	0.62	0.52	0.57	0.71
Physician Notes	0.96	0.96	0.96	0.96	0.99