

Introduction

Introduction

My background

- Clinical
 - Community health volunteer: Worked with various patient populations, e.g., prison population, LGBTQ+ patients, and also geriatric patients, providing HIV/HCV testing and counseling for over 5 years
- AI/NLP
 - Artificial Intelligence Scientist in Biotech/Pharma for 2 years (the vants) <- MLE at ESI <- speaker at Grace Hopper Celebration in 2018
 - Graduated from NYU with thesis MS in Biomedical Informatics in 2018 (*Zero Start: Deep Learning Models to Predict End of Life from Clinical Text*) and was contracted to deploy this work at various NYU Langone hospitals

Some topics you can expect from today:

Clinical NLP - unsupervised learning, text classification, named entity recognition, machine translation, various deep learning architectures (e.g., 1D CNNs, Seq2Seq, BERT-based and more!), data augmentation and generative type models, model deployment, clinical significance & collaboration (if time permits)

YEAR 2016

NIH Endowment Scholarship

The purpose of the scholarship is to provide financial support to high achieving, qualified students from socially or economically disadvantaged groups as defined by the National Institutions of Health (NIH).

I joined a top ten pharmacy program and was involved in many extracurriculars as well as wet lab research.





THE DEFINING

MOMENT

Deep Learning Models to Predict End of Life from Clinical Text

Isabel Kayu Metzger, MS , Seda Bilaloglu, MS , Vincent J Major, MS,
Himanshu Grover, PhD, Yindalon Aphinyanaphongs, MD/PhD
Department of Population Health, NYU Langone Health, New York, NY

Abstract

Accurate prognosis upon admission can help patients with serious diseases and their families to receive the care they need as they approach end of life. Admission notes hold key descriptions related to illness or problems and can be used to predict the course of the disease using machine learning techniques. In this paper, we explored deep

AMIA (2018)

Expanding the Reach of Structured EHR Data with Clinical Notes: Improving End-of-Life Prediction

Seda Bilaloglu, Vincent Major, Himanshu Grover, Isabel Kayu Metzger and
Yindalon Aphinyanaphongs
Department of Population Health, NYU Langone, New York, NY

Abstract

Appropriate treatment decisions and end-of-life planning for patients with serious, life-limiting diseases rely on accurate prognostic estimates. Many existing methods use unstructured electronic health record data which may limit generalizability across sites and restrict performance for patients with less documented history. Clinical notes may help to ‘level the playing field’. We use History and Physical (H&P) notes written within 16 hours of hospitalization to predict 60-day, all-cause mortality. We test several neural network approaches and observe little improvement over a CNN by adding bi-directional recurrence or convolutional attention. The CNN was prospectively validated against an existing system using structured data. The CNN reports

care planning, code status or advance directives, these patients may receive unwanted aggressive treatment. Precise identification of high-risk patients can break this cycle by encouraging appropriate end-of-life care.

Clinical risk tools often provide a score (Charlson et al. 1987; Knau et al. 1985; Morita et al. 1990) to stratify patients into risk groups. Numerous machine learning methods also exist but many are limited to specific populations by disease or acuity (Ghassemi et al. 2014; Makar et al. 2015; Elfify et al. 2017; Parikh et al. 2019). Several general approaches have been proposed for use to prompt clinicians to consider end-of-life planning (Avati et al. 2018; Wegier et al. 2019; Courtright et al. 2019; Major and Aphinyanaphongs 2020). Each of these works rely on structured electronic

Flairs Proceedings,
AAAI 2021

SMM4H Shared Task 2020 - A Hybrid Pipeline for Identifying Prescription Drug Abuse from Twitter: Machine Learning, Deep Learning, and Post-Processing

Isabel Metzger^{1,5}, Emir Y. Haskovic², Allison Black³, Whitley M. Yi⁴, Rajat S. Chandra¹, Mark T. Rutledge¹, William McMahon¹, and Yindalon Aphinyanaphongs⁵

¹Sumitovant Biopharma, Inc., New York, NY

²Lokavant, New York, NY

³Roviant Sciences, Inc., New York, NY

⁴Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado, CO

⁵Department of Population Health, NYU Langone Health, New York, NY

{isabel.metzger, rajat.chandra, mark.rutledge, bill.mcmahon} @sumitovant.com,

emir.haskovic@lokavant.com,

allison.black@roviant.com,

whitley.yi@ucdenver.edu,

yin.a@nyulangone.org

Association for Computational Linguistics, ACL (2020)

Question to SQL Query:

A Clinical Natural Language Processing Interface over Electronic Health Records Database

Isabel Metzger¹, Whitley Yi, PharmD²
New York University¹, NYU Langone Health², UNC Health Care²

ABSTRACT

Patient and clinician-generated narratives are considered “noisy text”, filled with domain-specific abbreviations and misspellings. Regardless, the presence of noise allows for knowledge discovery. When incorporated into machine learning based systems, along with structured clinical data, it has the potential to aid in clinical decision making and drive informed treatment. An interface over the electronic health record (EHR) that would translate questions from “natural language” to SQL queries, while accounting for domain-specific abbreviations, arises from realizing the need for a tool to aid clinicians in obtaining information for critical patient care decisions faster. For this reason, we built a prototype interface on the Medical Information Mart for Intensive Care III (MIMIC III) database, a publicly-available dataset of EHRs of critically ill patients from a composite via cross-sectional data of what is now nearly 4,000 questions from clinician assessments, and financial investigator perspectives) for data and statistics on interactions between patient demographics, treatments, comorbidities, and more. This dataset was manually extended by annotating the natural language questions with SQL which is used to query the MIMIC database. Medical concepts are recognized and normalized using a named entity recognition (NER) model pre-trained on requested annotated

METHODS

Named Entity Recognition Seq2Seq with Attention Connect to DB and Eval

How many patients died with <PROCEDURE_1> after receiving <PROCEDURE_1>?

```
select count(DISTINCT subject_id)
from PATIENTS
where expire_flag = 1 and subject_id IN
    (select distinct subject_id
     from DIAGNOSES_ICD
     where icd9_code IN
          (select distinct icd9_code
           from D_ICD_DIAGNOSES
           where long_title like 'PROBLEM_%')
     and subject_id IN
        (select distinct subject_id
         from PROCEDURES_ICD
         where icd9_code IN
              (select distinct icd9_code
```



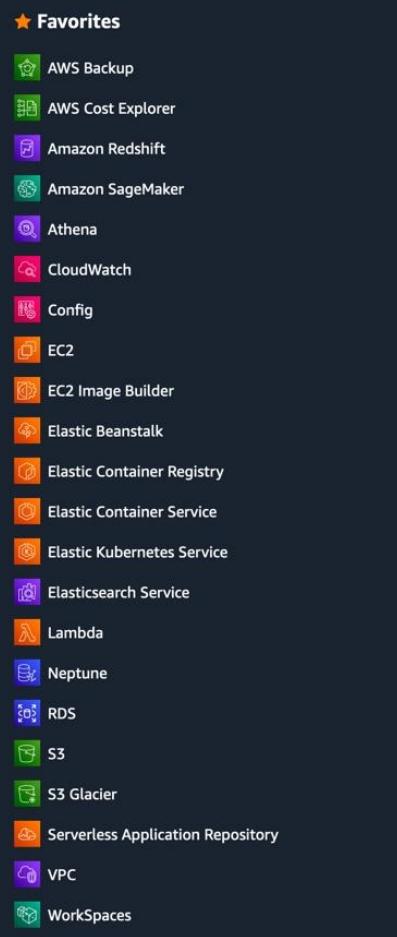
- RESULTS & CONCLUSION
- Bi-LSTM CRF on EHRs • Accuracy = 0.95
 - F-1 Score = 0.85
 - Bi-LSTM CRF on EHRs • Accuracy = 0.73
 - F-1 Score = 0.72
 - Seq2Seq with attention • Accuracy of 0.72
 - Accuracy of 0.72
- To understand the simulated experiments, the simulated experiments are then sequenced pairs of natural language

Northeast Health Summit hosted by
IBM & Brown University (2019)

ML Ops Experience

AWS Tech Stack

- Training either on Virtual Machines (EC2) instance and Sagemaker
- Dockerized models (ECR)
- Storage in S3 - Model, Data, Predictions
- Amazon Workspace
- Create API for my models so that the digital innovators could integrate them into their dashboards as well

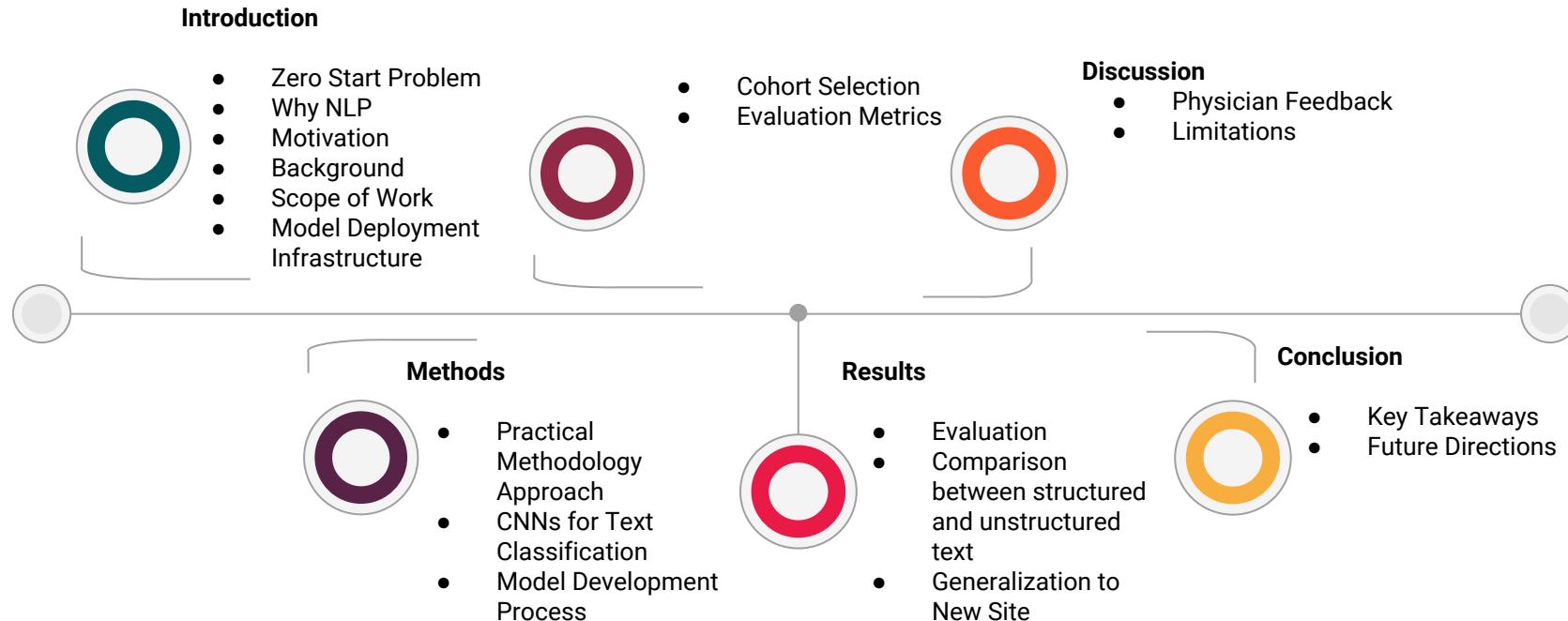


2017 - 2018

ZERO START:

Deep Learning Models to Predict Mortality from Clinical Text

Overview



ZERO START PROBLEM refers to when there are no previous records about a patient who visits us and thus we cannot use any data besides the note made upon admission to the facility. EPIC (healthcare software) alert system will not be triggered in these situations.

- **the goal of this work:** Can we predict 2-month mortality from unstructured data (clinical text), more specifically the History & Present illness (H&P) note?

Motivation

Why is this work important to do?

Why is this novel?

Most mortality predictive models fall into two extreme categories:

- Short term: (1-6 days) (e.g., rapid deterioration)
- Long term (12 months) (palliative care)

This work predicts for 2-6 months.

Why is this useful?

- Physician agreement with a recommendation for palliative care is highest at a prediction window of 2 months.
- Research shows that interventional palliative care is optimal when administered at least 6 months prior to mortality.

Improving palliative care with deep learning

Anand Avati,¹ Kenneth Jung,² Stephanie Harman,³ Lance Downing,² Andrew Ng,¹

Categories requested by physicians

Term	Definition
High Risk	Predicted to die within 2 months.
Appropriate	Expected to die within 6 monhts. GOC/ ACP warranted.
Inappropriate	Neither of the above.
Helpful	High Risk or Appropriate

Why NLP?

“Unstructured EHR” such as medical notes may provide unique insight and possibly more information than “structured”

A publication using the MIMIC III medical notes to **predict sepsis** within 24 hours found that the unstructured text data **performed better** than the structured tables. (Culliton, 2017)

We need to talk about death

Prognosis in Practice

- Clinicians are often inaccurate when predicting end-of-life (only right ~50% of the time). (Nicola White et. Al, 2016).

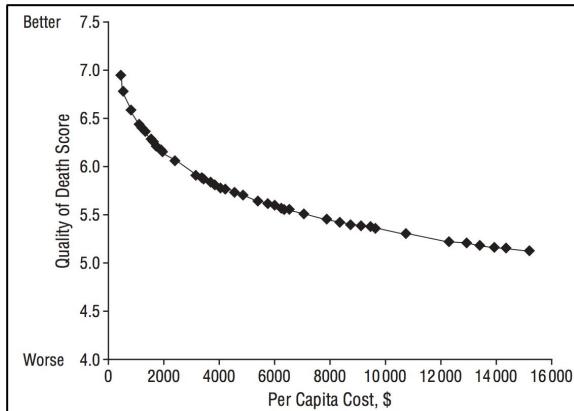
What is important to terminally ill patients?

- “Not to be kept alive on life support when there is little hope for a meaningful recovery” (55.7%)
- “That information about your disease be communicated to you by your doctor in an honest manner” (44.1%)
- “To complete things and prepare for life's end — life review, resolving conflicts, saying goodbye” (43.9%)

Costs

- A disproportionate amount of healthcare cost occurs in the last month of life (Zhang et al. 2009).
 - One study investigating the total cost of treatment in final week of life with, or without, an EOL conversation described:

higher costs were negatively associated with caregiver reported quality of death, and EOL conversations reduced total cost by 36% (\$2780 vs. \$1925).



(Zhang et al. 2009)

Score Based Metrics

The problem:

- Points-based metrics discretize real world physiology into coarse, weighted bins.
 - Makes them easy to use and interpretable
 - Restricted to integer weights, a small number of variables, a small number of bins

5 <0.17	3 0.17-4.94	Pre-ICU LOS 0 4.95-24.00 Hours	2 24.01-311.80	1 >311.80
		Age 0 <24 Years	3 24-53	6 54-77
10 3 - 7	4 8 - 13	GCS 0 15		9 78-89
		4 <33	Heart Rate 0 33-88 min ⁻¹	7 >90
4 <20.65	3 20.65-50.99	2 51-61.32	MAP 0 61.33-143.44 mmHg	1 89-106
		10 <6	3 >143.44	3 107-125
		1 6-12	Respiratory Rate 0 13-22 min ⁻¹	6 >125
3 <33.22	4 33.22-35.93	2 35.94-36.39	Temperature 0 36.40-36.88 °C	2 36.89-39.88
10 <671	5 671-1426.99	1 1427-2543.99	Urine Output 0 2544-6896 Cc/day	6 >39.88
			Ventilated 0 NO	8 >6896
		6 NO	Elective Surgery 0 YES	9 YES

Figure 1. Component weights and bins for the Oxford Acute Severity of Illness Score (OASIS). The **bold values** are the individual scores assigned to an associated range of measured values. For each variable, the worst score across the first day should be used to tabulate OASIS. The final OASIS score is the sum of all the component weights. LOS = length of stay, GCS = Glasgow Coma Score, MAP = mean arterial pressure.

(Johnson et al. 2013)

Scope of Work

SCOPE OF WORK

CENTER FOR HEALTHCARE INNOVATION AND DELIVERY SCIENCE

(CHIDS) of NYU LANGONE HEALTH is contracting with Isabel Metzger to build 2 month mortality

machine learning based models from initial history and physical notes of admitted patients. Accurate mortality prediction with initial notes would allow models to be built for patients that do not have prior data. These models help providers and patients coordinate in delivering supportive care that align with patient wishes. Specifically we will

- (1) integrate a model into our scalable text classification infrastructure,**
- (2) emit model classifications at a performance threshold daily to an inpatient team for feedback,**
- (3) deliver a publication ready writeup of the model and results.**

Model Deployment Infrastructure

PAU Operational/ Data Flow

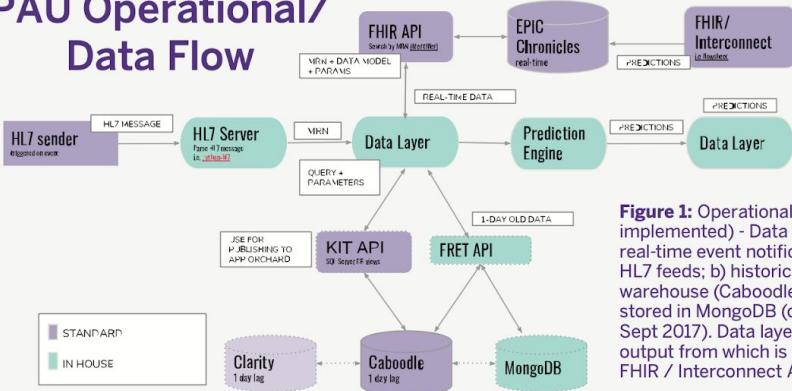


Figure 1: Operational data workflow (partially implemented) - Data infrastructure consumes a) real-time event notifications directly from Epic via HL7 feeds; b) historical data from Epic data warehouse (Caboodle) or Epic Clarity snapshot stored in MongoDB (currently has data from 2014-Sept 2017). Data layer feeds the prediction engine, output from which is populated back into Epic via FHIR / Interconnect API.

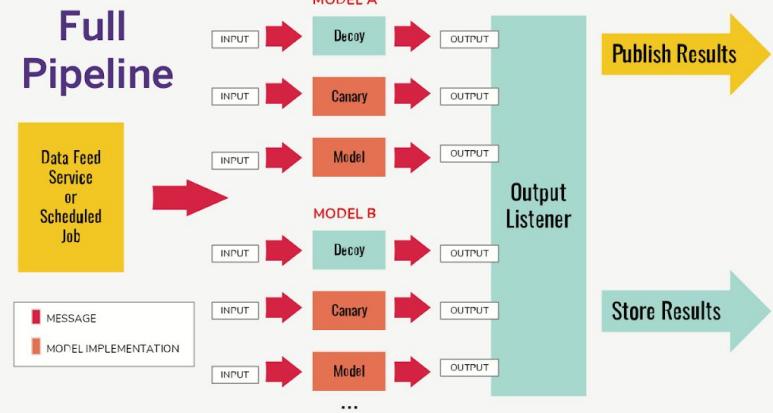


Figure 2: Production Model Deployment Architecture. Models implement a standardized interface and are deployed as containerized services that: a) consume input data stream; b) publish results for downstream services to listen and act on. Production models can be supplemented with Decoys (to capture raw input for later re-use) and Canaries (to monitor model drift in model or data). Streaming also facilitates comparing new test models that consume same input stream as production model.

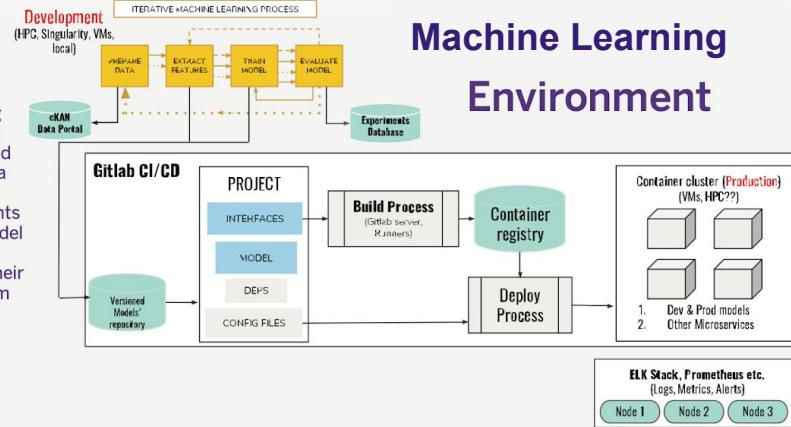


Figure 2: The Vision - A Dev cycle using High Performance Computing (HPC), Singularity, Virtual Machines (VMs) and local resources to (1) build datasets and store them (cKAN data portal), (2) build machine learning models, and (3) store the experiments (experiments database). Once a model is built, the Gitlab CI/CD allows versioning models, containerizing their deployment and finally pushing them into a Prod Container cluster for scalable deployment through microservices. Finally logs and monitoring are captured using ELK stack and Prometheus.

Practical Methodology

Practical Methodology

“Don’t Be a Hero - Best practices and literature review”

- Successfully applying deep learning requires more than just a good knowledge of what algorithms exist and the principles that explain how they work
- We also need to know how
 - to choose an algorithm for a particular application
 - to monitor and respond to feedback obtained from experiments in order to improve a machine learning system
- During development of deep learning systems, we need to decide:
 - whether to gather more data
 - increase or decrease model capacity
 - add or remove regularizing features
 - debug the software implementation of the model
- Understand what task you are solving what model architecture you should use and best practices

Task-specific best practices: Classification

- More so than for sequence tasks, where CNNs have only recently found application due to more efficient convolutional operations, CNNs have been popular for classification tasks in NLP.
- The following best practices relate to CNNs and capture some of their optimal hyperparameter choices.
- CNN filters: Combining filter sizes near the optimal filter size, e.g. (3,4,5) performs best (Kim, 2014; Kim et al., 2016). The optimal number of feature maps is in the range of 50-600 (Zhang & Wallace, 2015) .

Aggregation function: 1-max-pooling outperforms average-pooling and k-max pooling (2015)

Sebastian Ruder, "Deep Learning for NLP Best Practices".

<http://ruder.io/deep-learning-nlp-best-practices/>

Model Development Process

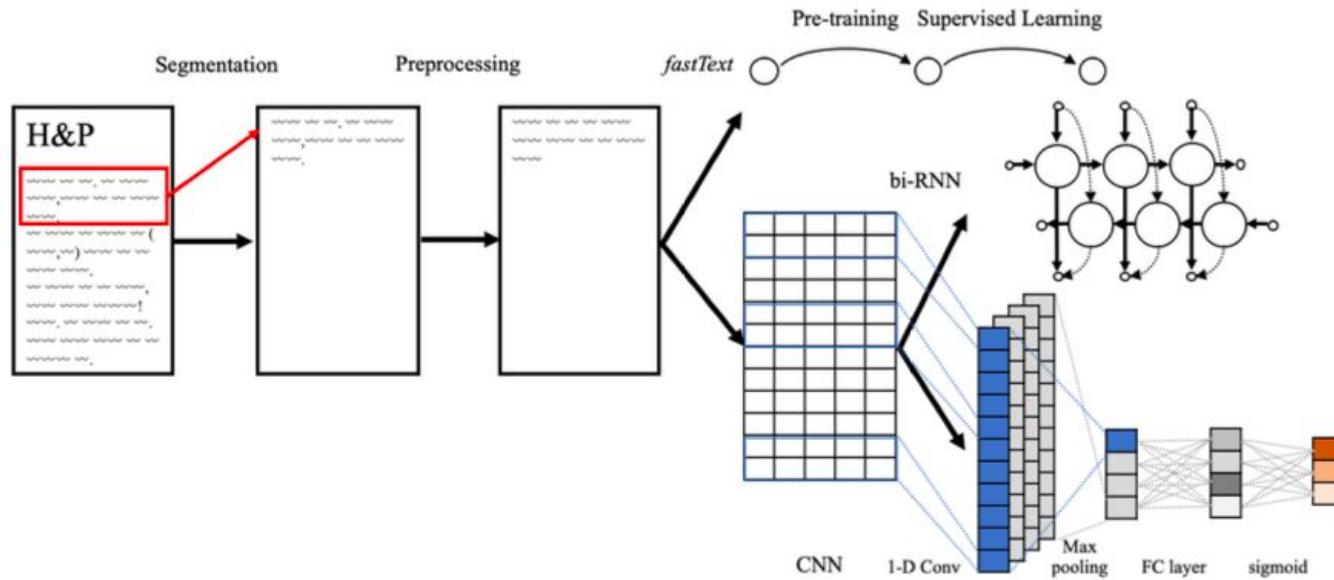


Figure 1: Preprocessing and model development workflow

CNNs in Natural Language Processing

- Character/Byte level, e.g., SMILES CC1=C(C=C(C=C1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]
- To reduce the vocabulary size
- Orthology
- **When you want things to run faster *and significantly less computationally expensive***

Convolutional Neural Networks for Sentence Classification (Kim, 2014)

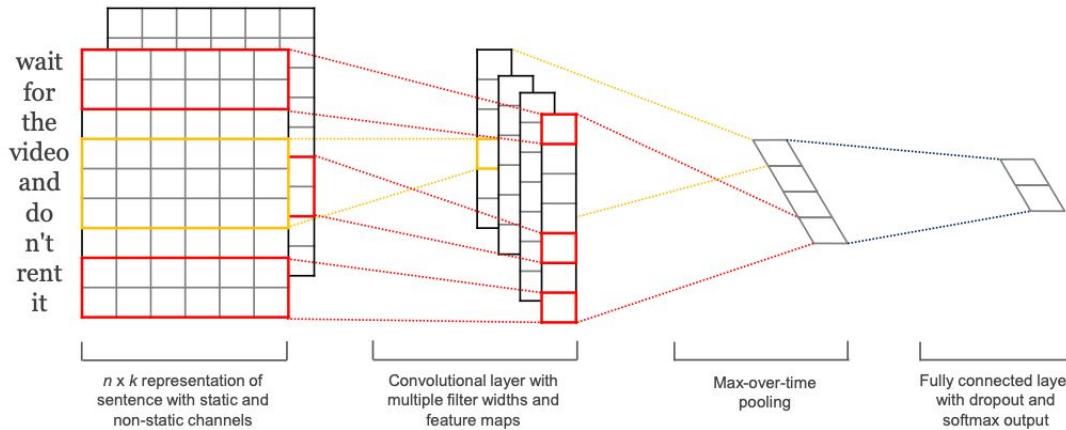
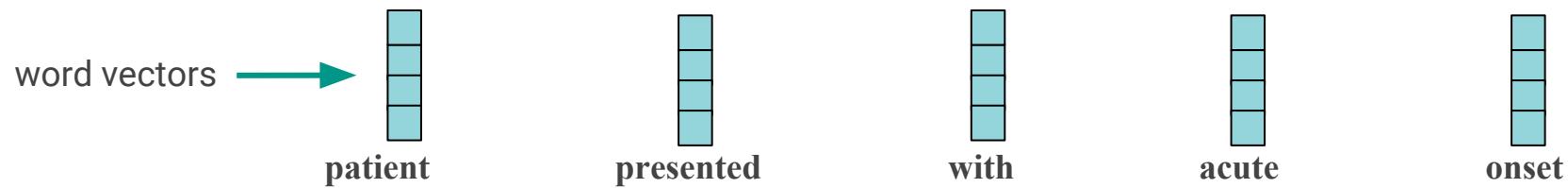


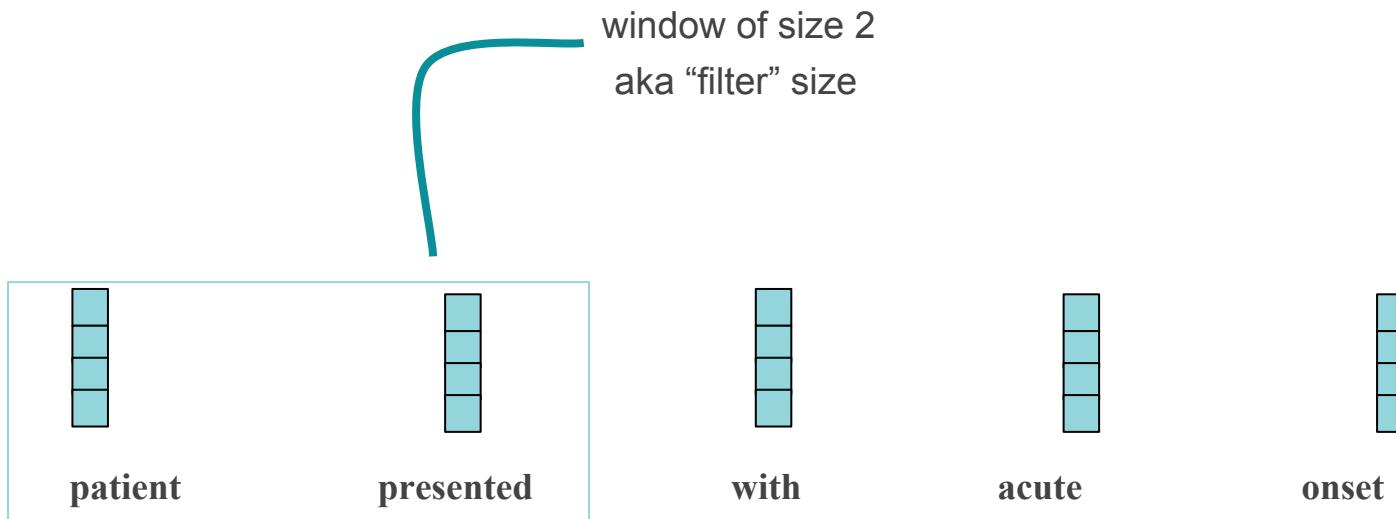
Figure 1: Model architecture with two channels for an example sentence.

How to convolve on text



"channel" size is 4 in this example

How to convolve on text



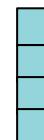
How to convolve on text



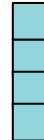
patient presented



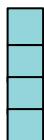
patient



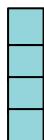
presented



with



acute



onset

How to convolve on text



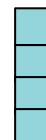
patient presented



presented with



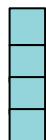
patient



presented



with

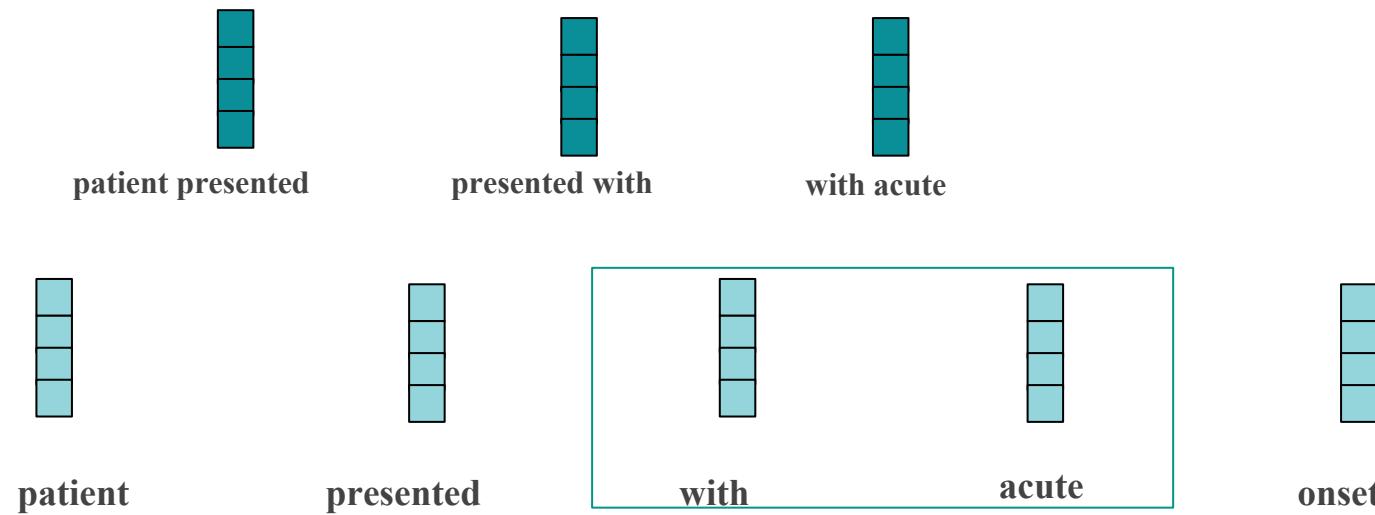


acute



onset

How to convolve on text



How to convolve on text



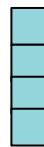
patient presented



presented with



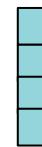
with acute



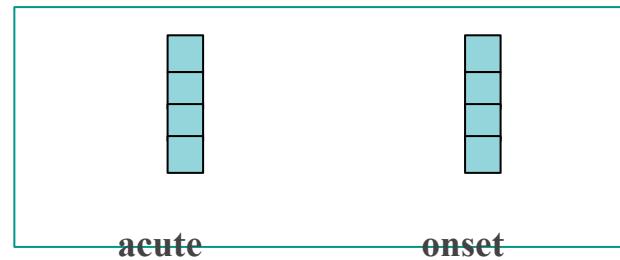
patient



presented



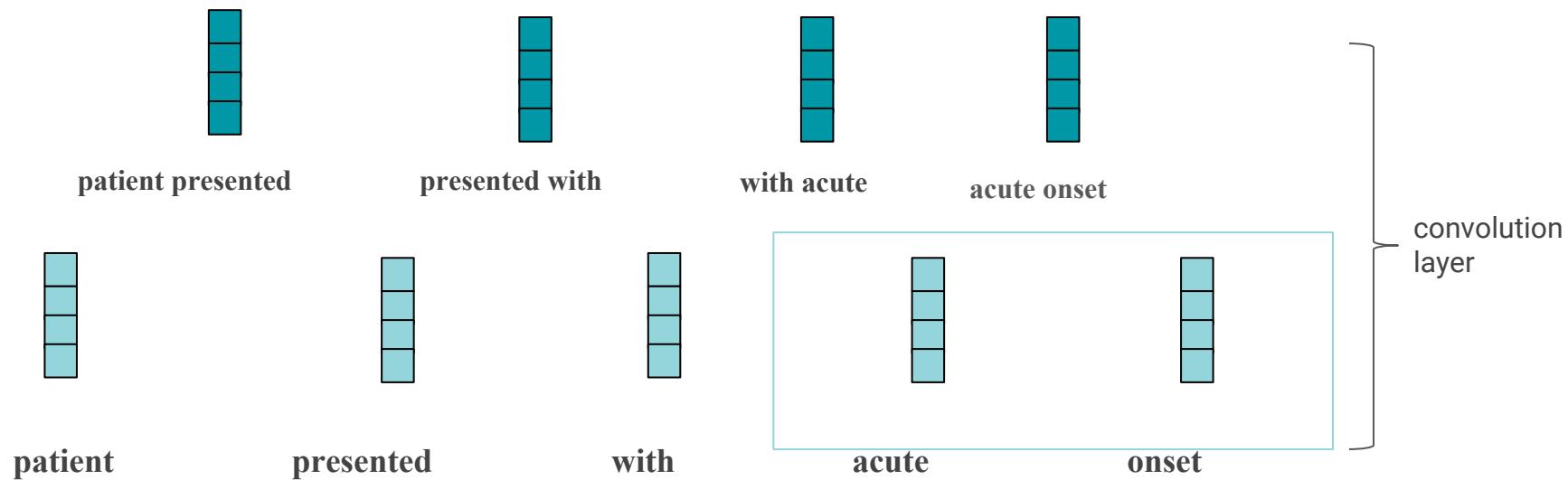
with



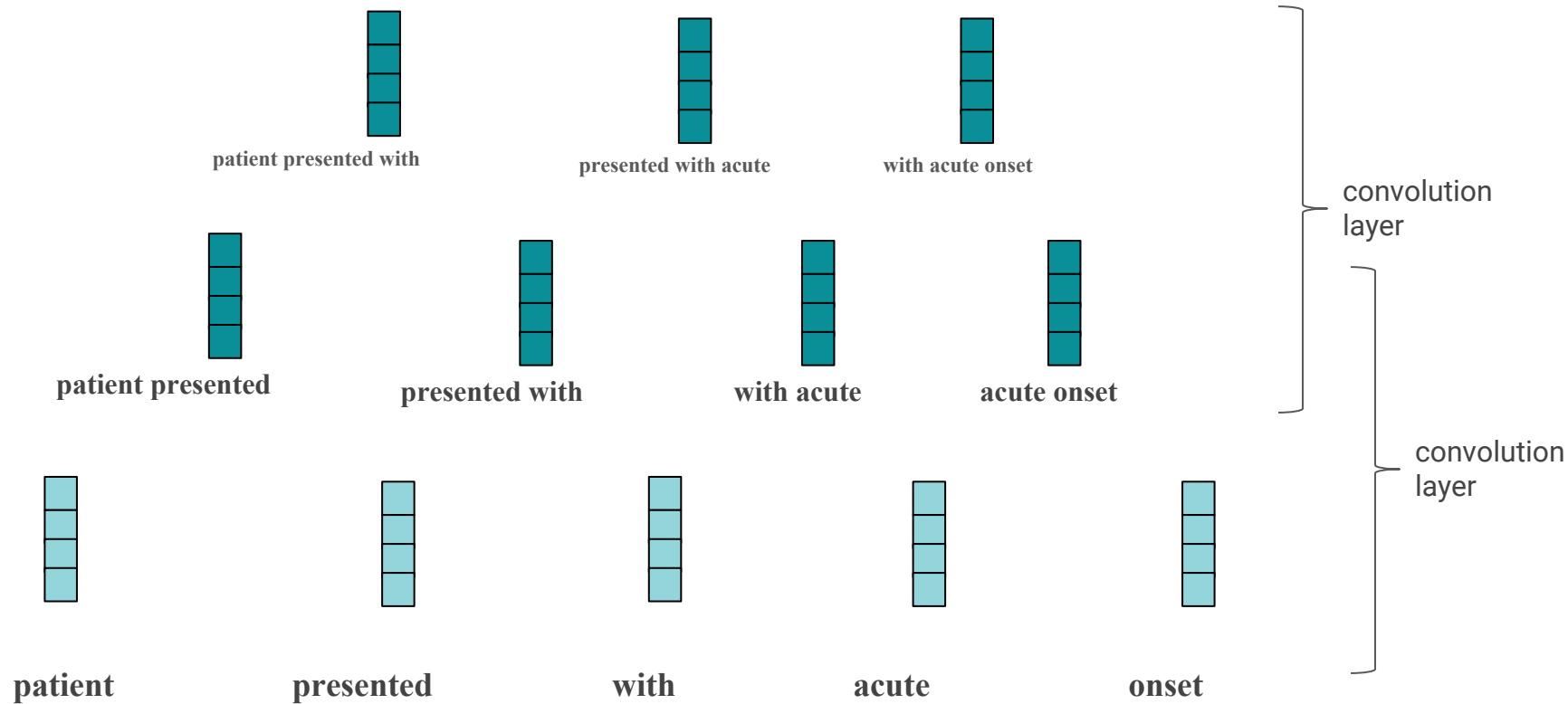
acute

onset

How to convolve on text



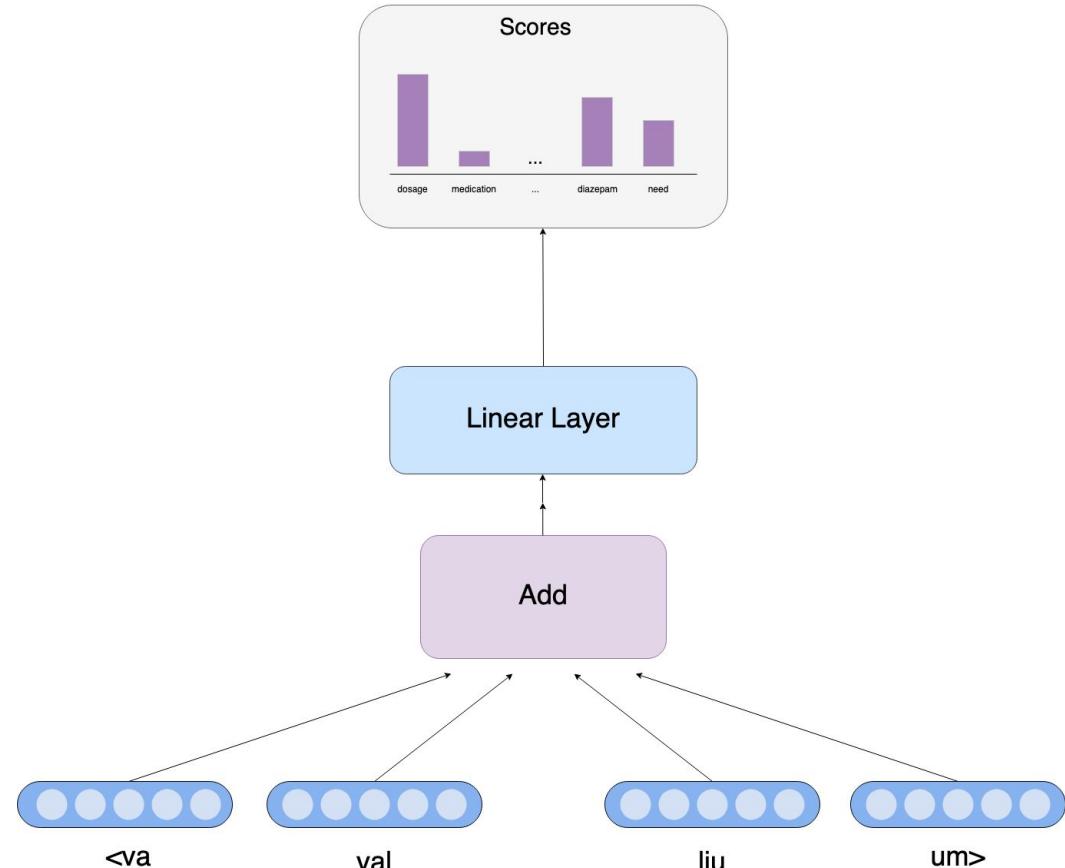
How to convolve on text



Unsupervised Learning

Word Embeddings with fastText

- Word2Vec < Glove < fastText
- 100 and 300 dim skip-gram model with negative sampling and subword enrichment
- training set H&P notes (62k documents and 5M words) + all MIMIC-III Critical Care Database (Johnson et. al. 2016)(2M documents and 5M Words)



Qualitative Evaluation of Word Embeddings

Testing clinical domain-specific acronym against facebook's fastText pretrained embeddings

(300 dim) MIMIC III
+ NYU embeddings

fastText's English
Wikipedia 300 dim

Query word? **dnr**

dni 0.942678
hcp 0.749637
resuscitate 0.711382
code 0.65287
intubate 0.638306
cmo 0.637459
dnri 0.612248
reintubate 0.606106
hcps 0.603236
hospice 0.58464

Query word? **dnr**

dnl 0.954677
dnssec 0.940462
cwp 0.936214
dpb 0.934438
hvdc 0.929971
bvu 0.927638
dnq 0.927279
tpb 0.925606
pkc 0.924558
hvd 0.923596

DNR = do not resuscitate

DNI = do not intubate

HCP = health care provide

CMO = comfort measures only

Cohort Selection

- Restricted to notes between 0-16 hours upon admission
- Notes with more than 50 words (after removing addendums and attestations)
- Restricted to most common author types: Physician, Fellow, Resident, Physician Assistant, and Nurse Practitioner

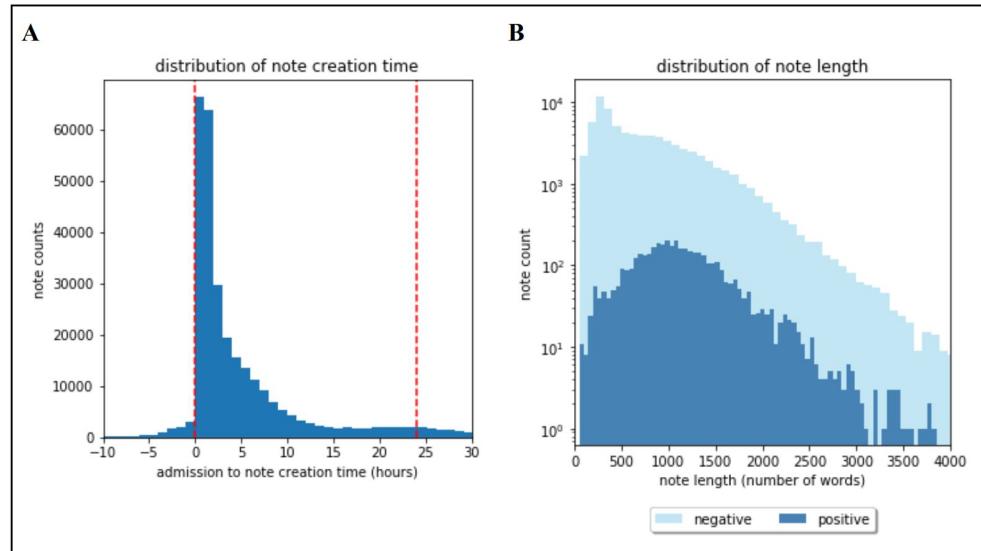


Figure 1. A) Time taken between admission and note creation, B) distribution of note length

Cohort Selection

- Hospital admissions from January 1st, 2013 to December 31st, 2017 (5 years)
- Death outcomes from social security data and institutional data (positive labels)
- Separation performed temporarily and at patient level to prevent data leakage between sets (Neto et. al. 2019)

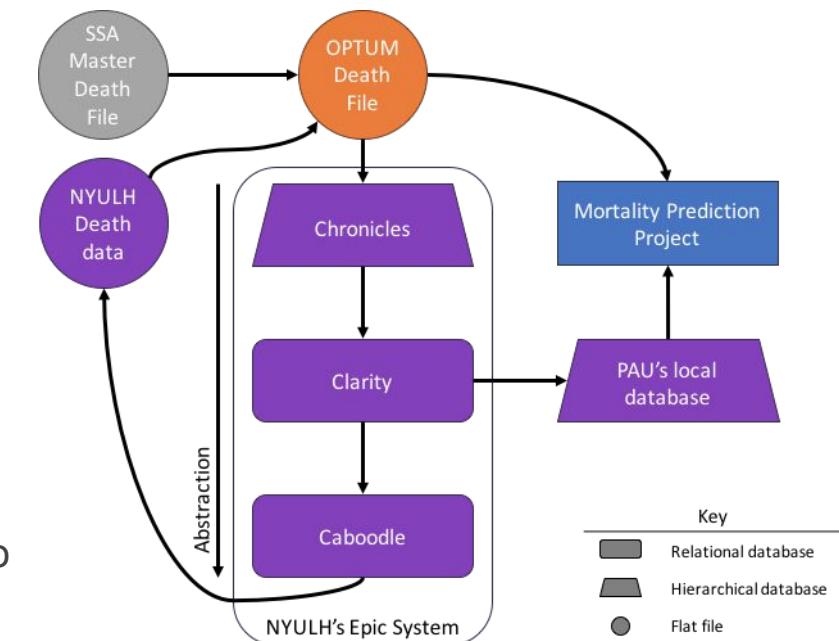


Table 1: Datasets of H&P notes and 60-day patient outcomes used for model development.

	Training	Validation	Test	Total	
Positive	2,679	468	1,055	4,202	(5.1%)
Negative	50,140	8,973	19,473	78,586	
Total	52,819	9,441	20,528	82,788	

Results

Evaluation Metrics

- End-of-life is a rare outcome which can skew evaluation metrics such as accuracy
- Visualized with receiver operating characteristic (ROC) and measured by the AUROC
- As this model is potentially helpful to recommend an intervention to predicted positives, precision-recall curves (PRC) and AUPRC and the max-F1 score are employed

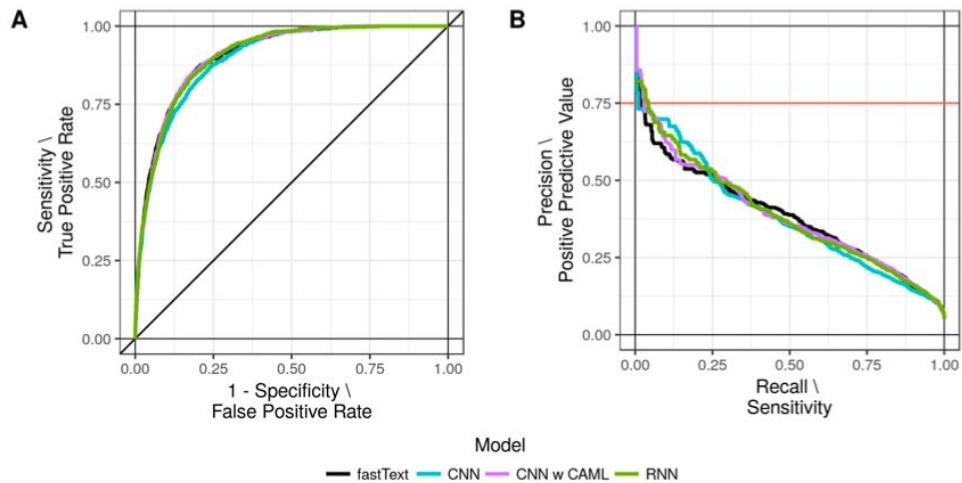


Figure 2: Test set evaluation metrics. A) ROC and B) PRC.

Table 2: Test set evaluation metrics.

Model	AUROC [95% CI]	AUPRC [95% CI]	max F-1 [95% CI]
CNN	0.899 [0.890, 0.908]	0.381 [0.348, 0.421]	0.418 [0.394, 0.450]
RNN (bi-GRU)	0.907 [0.899, 0.915]	0.388 [0.357, 0.427]	0.421 [0.399, 0.452]
CNN with CAML	0.908 [0.900, 0.917]	0.388 [0.354, 0.424]	0.425 [0.401, 0.454]

CNNs trained & predicted 75% faster than RNN with comparable results, thus this model was selected for prospective validation

Performance on New Site (New Paper)

Evaluation of the model on an entirely different hospital

Table 3: Prospective validation results.

Metric	H&P CNN	Structured Data
Total admissions		65,727
Predictions	57,997	53,446
Admissions predicted	37,720 (57.4%)	53,446 (81.3%)
Timing (hrs)	29.6	0.03
median [IQR]	[19.0, 36.8]	[0.02, 0.85]
High-risk admissions	80 (0.21%)	131 (0.25%)
AUROC	0.860	0.806
[95% CI]	[0.847, 0.873]	[0.793, 0.820]
AUPRC	0.314	0.179
[95% CI]	[0.282, 0.353]	[0.157, 0.204]
Max F-1	0.377	0.228
[95% CI]	[0.352, 0.409]	[0.210, 0.254]

- Identified 26 high-risk patients using H&P approach vs 1 from the structured data model
- The text-based approach outperforms an structured data system and generalizes better to a new hospital location

Comparison of the models that performed on a new unseen 4th hospital (8 months)

Why this is relevant to Covera Health?

- Although notes can be written differently the model still generalizes quite well
- Note the same for the model built on structured data
 - Not necessarily the same for hospital - billing behavior are often dictated by the contract with payers

Clinical Significance and Feedback

Feedback from Clinicians

“One case (I can provide the details if that would be helpful), at the beginning of the admission to me would not have triggered that pt was an end of life patient, but as the admission unfolded this became clear as pt became more acutely ill. Since you are running this tool retrospectively I am curious if this patient would have flagged at the start of pt admission.”

① Mortality Predictor

This patient has been identified as high risk for dying in the next two months. This notification will be presented to the unit medical director and chief of service. Within the clinical context of this presentation please consider:

1. The overall care trajectory and the impact of any intervention within that context
2. The identified opportunity for an advanced care planning conversation during this admission
3. Consulting palliative care or geriatrics if you have not done so already

I agree with the above:

Order

Do Not Order

Mandatory Surprise Question

The following actions have been applied:

Sent: This advisory has been sent via In Basket

② Acknowledge Reason

I do not agree with the above

I need to further assess the patient

Accept

- Presented to the attending of record
- Interruptive (cannot do away)
 - *One chance to further assess the patient*

Prompts a MSQ (mandatory surprise question) order and suggests an ACP conversation and consulting Palliative Care/Geriatrics

References

- Avati A, Jung K, Harman S, et al. (2017) Improving Palliative Care with Deep Learning. arXiv [cs.CY]. Available from: <http://arxiv.org/abs/1711.06402>.
- Choi E, Bahadori MT, Schuetz A, et al. (2016) Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. JMLR workshop and conference proceedings 56: 301–318.
- Detsky ME, Harhay MO, Bayard DF, et al. (2017) Discriminative Accuracy of Physician and Nurse Predictions for Survival and Functional Outcomes 6 Months After an ICU Admission. *JAMA: the journal of the American Medical Association* 317(21): 2187–2195.
- Detsky ME, Harhay MO, Bayard DF, et al. (2017) Discriminative Accuracy of Physician and Nurse Predictions for Survival and Functional Outcomes 6 Months After an ICU Admission. *JAMA: the journal of the American Medical Association* 317(21): 2187–2195.
- Ghassemi MM, Richter SE, Eche IM, et al. (2014) A data-driven approach to optimized medication dosing: a focus on heparin. *Intensive care medicine* 40(9): 1332–1338.
- Huang J, Osorio C, Sy LW. (2018) An Empirical Evaluation of Deep Learning for ICD-9 Code Assignment using MIMIC-III Clinical Notes. arXiv [cs.CL]. Available from: <https://arxiv.org/abs/1802.02311>.
- Johnson AEW, Pollard TJ, Shen L, et al. (2016) MIMIC-III, a freely accessible critical care database. *Scientific data* 3: 160035.
- Li J, Chen X, Hovy E, et al. (2016) Visualizing and Understanding Neural Models in NLP. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 681–691.
- Longo D, Fauci A, Kasper D, et al. (2011) Harrison's Principles of Internal Medicine, 18th Edition. McGraw Hill Professional.
- Konkle B. Chapter 115. Disorders of Platelets and Vessel Wall. In: Longo DL, Kasper DL, Jameson JL, Fauci AS, Hauser SL, eds. *Harrison's Principles of Internal Medicine*. 18th ed. New York: McGraw-Hill; 2012.<http://www.accesspharmacy.com/content.aspx?aid=9100733>. Accessed January 30, 2013.
- Mullenbach J, Wiegreffe S, Duke J, et al. (2018) Explainable Prediction of Medical Codes from Clinical Text. arXiv [cs.CL]. Available from: <https://arxiv.org/abs/1802.05660>.
- Purushotham S, Meng C, Che Z, et al. (2017) Benchmark of Deep Learning Models on Large Healthcare MIMIC Datasets. arXiv [cs.LG]. Available from: <http://arxiv.org/abs/1710.08531>.
- Smythe MA, Prizziola J, Dobesh PP, et al. (2016) Guidance for the practical management of the heparin anticoagulants in the treatment of venous thromboembolism. *Journal of thrombosis and thrombolysis* 41(1): 165–186.

AI Scientist - Journey into Vant Alliance



Rovant Sciences

Sumitomo Dainippon Pharma and Rovant Sciences Enter into a Memorandum of Understanding to Create Broad Strategic Alliance to Deliver Promising New Medicines to Patients

- Sumitomo Dainippon-Rovant Alliance ("Alliance") encompasses up to 11 biopharmaceutical Vants with more than 25 innovative clinical programs and multiple potential product launches from 2020 to 2022, and access to key elements of Rovant's proprietary technology platforms including DrugOme and Digital Innovation

- Sumitomo Dainippon Pharma to enter into contract agreements with Rovant Health technology Vants including Datavant and Alyvant

- Sumitomo Dainippon Pharma to take over 10% equity stake in Rovant

- Parties have entered into a non-binding memorandum of understanding ("Memorandum"); a definitive agreement expected by the end of October 2019

NEWS PROVIDED BY
Rovant Sciences; Sumitomo Dainippon Pharma Co., Ltd. →
Sep 05, 2019, 22:40 ET

SHARE THIS ARTICLE

TOKYO, OSAKA, Japan, LONDON, and BASEL, Switzerland, Sept. 5, 2019 /PRNewswire/ -- Sumitomo Dainippon Pharma Co., Ltd. (TSE: 4506), a leading Japanese pharmaceutical company, and Rovant Sciences Ltd., a technology-enabled healthcare company, today announced that they have entered into the Memorandum for the creation of a novel and broad Alliance to include the transfer to Sumitomo Dainippon Pharma of Rovant's ownership interests in 5 of their biopharmaceutical companies ("Vants"), with options to acquire up to 6 additional Vants, and access to Rovant's proprietary technology platforms, DrugOme and Digital Innovation. Rovant will collaborate with Sumitomo Dainippon Pharma with the continued involvement of Rovant's senior leaders to ensure the success of the Alliance. In addition, Sumitomo Dainippon Pharma will take an equity stake of over 10% of shares outstanding in Rovant.



Sumitovant

"The Alliance"

Generating Toxic Molecules

- Authors: Izzy Metzger Computational Research, Zach Carpenter Roivant Health
 - paper we follow is: <https://arxiv.org/abs/1610.02415>
 - other papers we utilize: <https://link.springer.com/article/10.1007%2FBF00332918>
 - <https://github.com/microsoft/molecule-autencoder>
 - Note: deepchem has a vae model that follows the Aspuru-Guzuki Framework similarly to ours but we didn't implement deepchem in this (we could in the future if everyone prefers that framework)
 - we trained the model using the latest chembl database (chembl 25)
 - This notebook shows the output of that model
 - we will use the same latent dimension as in the paper (292)
 - OVERALL DESIGN:
 - From each toxic smile in the dataset sample the latent space next to that toxic smile to get auto-generated toxicish molecules
 - In particular, we create 1000 new toxicish smiles using this method with our vae model (that we trained over the weekend on the latest chembl db)
 - Each of those generated smiles are then checked to see if they are "working" via rdkit/ and get rid of the broken smiles (for e.g., 1000 generated toxicish smiles in this example (which uses the first toxic (label==1) smile in the latest master table)
 - We plot valid smiles (note some of these are real and already exist and some of them are not real)
 - we can also interpolate two smiles (like in the paper-- note our results are not as good as the paper but we didn't train the model as long)
- NEXT STEPS:
- Determine measures of significance
 - Defining a toxic metric for our new smiles
 - Some we can validate, others we can look at tanimoto similarity
 - <https://stackoverflow.com/questions/51681659/how-to-use-rdkit-to-calculate-molecular-fingerprint-and-similarity-of-a-list-of-smiles>
 - Tanimoto method is then used to produce a dissimilarity matrix related to the Jaccard dissimilarity
 - Tanimoto dissimilarity is a modified Hamming dissimilarity
 - Perhaps using LIME in some way?
 - The paper posted describes how we could optimize a molecule produced by our model for a particular molecular property (e.g., skin permeability or even solubility .. possibility toxicity?)

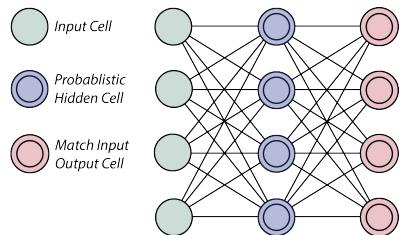
Zach Carpenter now the CEO of Vant.AI



name	canonical	isomeric	toxicity	one_hot	designation	Mol_ID	ABC	...	SRM
S=C=Nc1c2c(ccc1)cccc2	C1=CC=C2C(=C1)C=CC=C2N=C=S	C1=CC=C2C(=C1)C=CC=C2N=C=S	Hepatotoxicity	1	experimental/Liu smiles	Mol0	9.818615	...	9.3825
c1(c(cc1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[...]	CC1=C(C=C(C=C1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])	CC1=C(C=C(C=C1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])	Hepatotoxicity	1	experimental/Liu smiles	Mol1	11.877237	...	9.6371
c1(c(cc1[N+](=O)[O-])[N+](=O)[O-])O	C1=CC(=C(C=C1[N+](=O)[O-])[N+](=O)[O-])O	C1=CC(=C(C=C1[N+](=O)[O-])[N+](=O)[O-])O	Hepatotoxicity	1	experimental/Liu smiles	Mol2	9.618017	...	9.3008
O(CCO)CC	CCOCOC	CCOCOC	Hepatotoxicity	1	experimental/Liu smiles	Mol3	3.535534	...	6.6080
Oc1cc2c(cc1)cccc2	C1=CC=C2C=C(C=CC2=C1)O	C1=CC=C2C=C(C=CC2=C1)O	Hepatotoxicity	1	experimental/Liu smiles	Mol4	8.554231	...	9.2251

Variational AutoEncoders

CCC(O)(\C=C\Cl)C



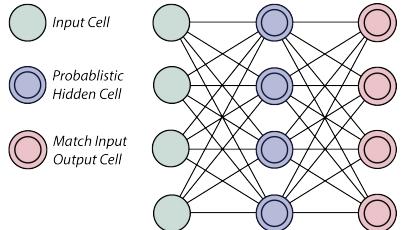
```
array([[ 0.05363178,  0.0078782 , -0.07700312,  0.05649856,  0.13152978,
       0.07801478, -0.07673378, -0.04488152,  0.04152625,  0.05349444,
      -0.04960863,  0.00735946, -0.03721184,  0.06771883, -0.00867662,
     -0.11634302, -0.07169832, -0.12267046,  0.06691165,  0.06741397,
      0.05132991, -0.11788134, -0.04044838, -0.06816725,  0.02068123,
      0.05099233,  0.18753207, -0.13674015, -0.12335098,  0.0800745 ,
      0.05587533,  0.10807797, -0.03181177, -0.1161322 , -0.138663 ,
     -0.00987642,  0.00706142,  0.19610442,  0.0629833 , -0.05595475,
     0.01440607,  0.01905868, -0.07988203,  0.05325389,  0.04673581,
    -0.13123605,  0.04979331,  0.10067537,  0.11553518,  0.03937402,
     0.01276392,  0.17929122,  0.08065402,  0.0651684 , -0.11405068,
     0.12338633, -0.09664553, -0.04511814, -0.05998807,  0.10256029,
     0.10905663, -0.06486235, -0.12057194, -0.08205813, -0.0080945 ,
    -0.00697902, -0.14520209,  0.00825424,  0.01416555, -0.10578428,
    -0.01481704, -0.11556359, -0.03562332, -0.04776421, -0.03348266,
    0.13707504,  0.16890147,  0.14823346,  0.12322275, -0.03067702,
     0.04986928, -0.00574416, -0.02069437, -0.01018216,  0.0542658 ,
     0.02804135, -0.05027196,  0.08862189,  0.0130431 , -0.03631146,
     0.01536888,  0.05524712,  0.08686657,  0.04180029,  0.0685646 ,
    -0.13962124,  0.00456896,  0.0190599 ,  0.20133126,  0.12001671,
     0.01564296, -0.02090459,  0.07305951,  0.00762762,  0.08478145,
    -0.01544177, -0.03805823,  0.05480809,  0.02945027,  0.04178546,
    -0.0258328 , -0.06435941,  0.04975348,  0.00826132, -0.0732389 ,
     0.12626693,  0.03737481, -0.00254592,  0.03650747,  0.01470203,
     0.01745417,  0.01659185, -0.01863235, -0.09864108, -0.02229385,
     0.02127865, -0.08553039, -0.01834461,  0.02821246,  0.17647676,
     0.05512521,  0.00616604, -0.04119888,  0.02849228, -0.04119967,
    -0.01906603, -0.02676078, -0.01619593, -0.04449657, -0.06778533,
     0.01248086,  0.13863096,  0.09380561,  0.01931311,  0.01201088,
     0.07519506, -0.04049241,  0.00817274, -0.12936787,  0.10033616,
```



- Converting SMILE to latent representation to molecule and Vice Versa

Variational AutoEncoders

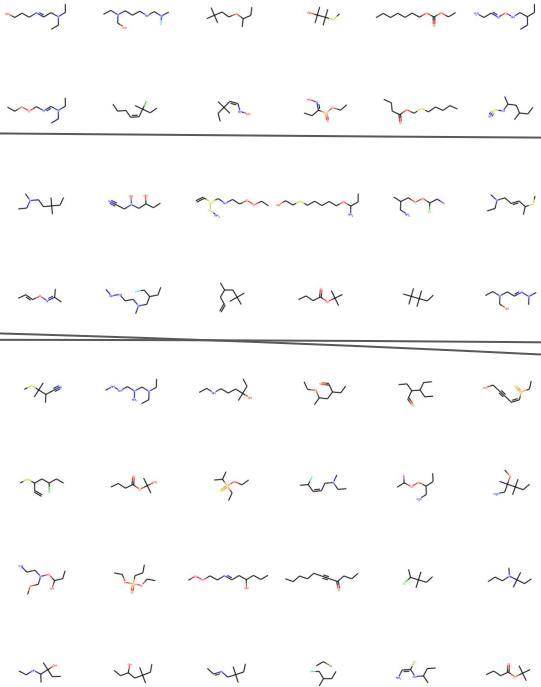
CCC(O)(\C=C\Cl)C



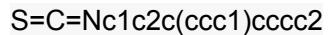
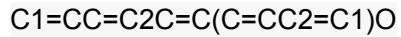
```
array([[ 0.05363178,  0.0078782 , -0.07700312,  0.05649856,  0.13152978,
       0.07801478, -0.07673378, -0.04488152,  0.04152625,  0.05349444,
      -0.04960863,  0.00735946, -0.03721184,  0.06771883, -0.00867662,
     -0.11634302, -0.07169832, -0.12267046,  0.06691165,  0.06741397,
      0.05132991, -0.11788134, -0.04044838, -0.06816725,  0.02068123,
      0.05099233,  0.18753207, -0.13674015, -0.12235098,  0.0800745 ,
      0.05587533,  0.10807797, -0.03181177, -0.1161322 , -0.138663 ,
     -0.00987642,  0.00706142,  0.19610442,  0.0629833 , -0.05595475,
     0.01440607,  0.01905868, -0.07988203,  0.05325389,  0.04673581,
    -0.13123605,  0.04979331,  0.10067537,  0.11553518,  0.03937402,
     0.01276392,  0.17929122,  0.08065402,  0.0651684 , -0.11405068,
     0.12338633, -0.09664553, -0.04511814, -0.05998807,  0.10256029,
     0.10905663, -0.06486235, -0.12057194, -0.08205813, -0.0080945 ,
    -0.00697902, -0.14520209,  0.00825424,  0.01416555, -0.10578428,
    -0.01481704, -0.11556359, -0.03562332, -0.04776421, -0.03348266,
     0.13707504,  0.16890147,  0.14823346,  0.12322275, -0.03067702,
     0.04986928, -0.00574416, -0.02069437, -0.01018216,  0.0542658 ,
     0.02804135, -0.05027196,  0.08862189,  0.0130431 , -0.03631146,
     0.01536888,  0.05524712,  0.08686657,  0.04180029,  0.0685646 ,
    -0.13962124,  0.00456896,  0.0190599 ,  0.20133126,  0.12001671,
     0.01564296, -0.02090459,  0.07305951,  0.00762762,  0.08478145,
    -0.01544177, -0.03805823,  0.05480809,  0.02945027,  0.04178546,
    -0.0258328 , -0.06435941,  0.04975348,  0.00826132, -0.0732389 ,
     0.12626693,  0.03737481, -0.00254592,  0.03650747,  0.01470203,
     0.01745417,  0.01659185, -0.01863235, -0.09864108, -0.02229385,
     0.02127865, -0.08553039, -0.01834461,  0.02821246,  0.17647676,
     0.05512521,  0.00616604, -0.04119888,  0.02849228, -0.04119967,
    -0.01906603, -0.02676078, -0.01619593, -0.04449657, -0.06778533,
     0.01248086,  0.13863096,  0.09380561,  0.01931311,  0.01201088,
     0.07519506, -0.04049241,  0.00817274, -0.12936787,  0.10033616,
```

Looking for 1k toxic neighbours

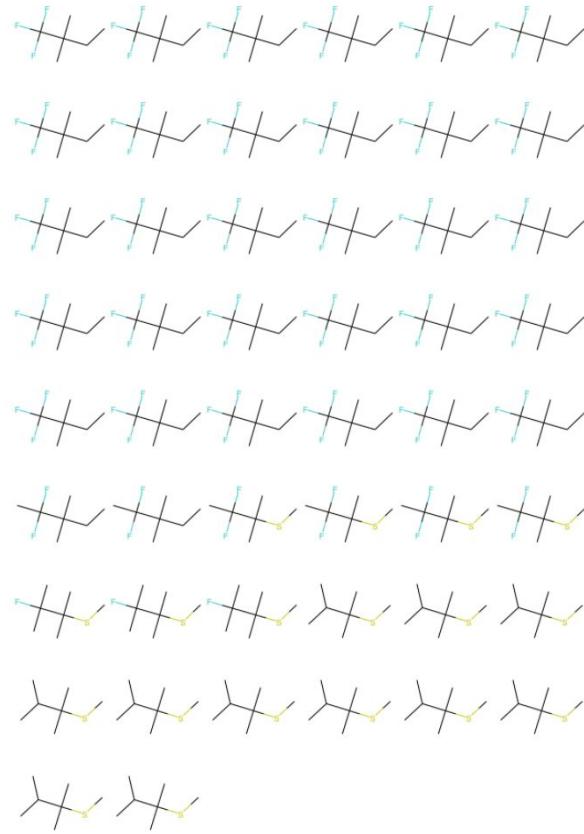
- Generating 1k random 292 dimension continuous arrays with mean=tox_latent and stdev==0.1 to sample the latent space next to the tox example.



VAE



- **Interpolation of two toxic molecules**
- **(combinations can be endless!)**



30+ models I developed during my two years in Computational Research Team (DrugOME) in areas (1 Natural Language Processing & 2 Drug Development)

DrugOME™

The DrugOME harnesses the power of known data to inform decision-making across the entire drug development continuum

Just as the genome comprises all data related to gene structure and expression and the proteome comprises all data related to the protein structure, location, function and interaction, Sumitovant's DrugOME comprises a vast amount of data related to drug development

Sumitovant's DrugOME integrates three powerful realms to provide unparalleled insights into the potential challenges and opportunities of specific molecules and drug formulations in specific clinical indications and lines of therapy

1
Natural Language Processing

Natural language processing realm uses automated systems to explore published literature, texts, documents and news to identify potential assets for acquisition/licensing; discover novel science; and identify key opinion leaders who can provide critical insight and champion new product opportunities

2
Drug Development

Drug Development realm examines potential drug molecules, drug targets, clinical trial data and drug development companies to identify and value assets; predict toxicity; define the competitive and therapeutic landscape for a specific asset; predict clinical trial costs; enable repositioning of existing assets into new and valuable indications; support high-value partnerships and collaborations

3
Real-world Data and Evidence

Real-world data and evidence realm utilizes patient, physician and payer data to support more accurate and effective market characterization and product marketing efforts; identify existing and evolving trends in treatment patterns, treatment costs and epidemiology; provide insight into the patient journey and potential barriers to adoption of or compliance with particular therapies; optimize clinical trial site and investigator selection; and enable virtual clinical trials

Language Modelling & Transformer-Based Projects

- Reading Comprehension (Question and Answering)
- Textual Entailment within Pharma Patents, Regulatory Documents and More
- PICO (Patient, Intervention, Cohort, and Outcome) NER and Relationship Extraction
- Medication, Dosage, etc NER & Clinical Relationship Extraction

Transformers Tokenization

```
# bert-base-uncased
```

```
['[CLS]', 'laced', 'with', 'dreams', '-', 'dripping', 'in', 'reality', ',', 'the', 'american',  
'dream', 'reign', '#ites', 'after', '9', '.', '11', 'with', 'a', 'true', 'story', 'about',  
'the', 'devil', 'ray', "!", 's', 'mid', '-', 'life', 'rookie', ',', 'jimmy', 'morris', '.',  
'[SEP'] ]
```

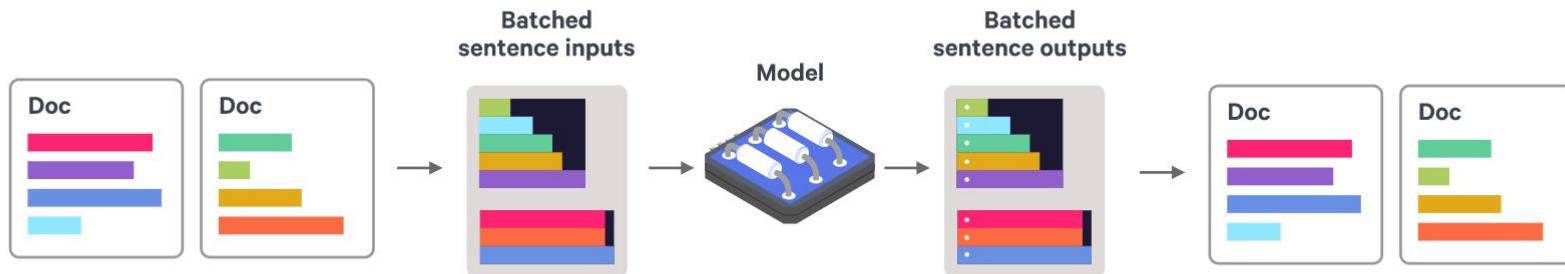
```
# gpt2
```

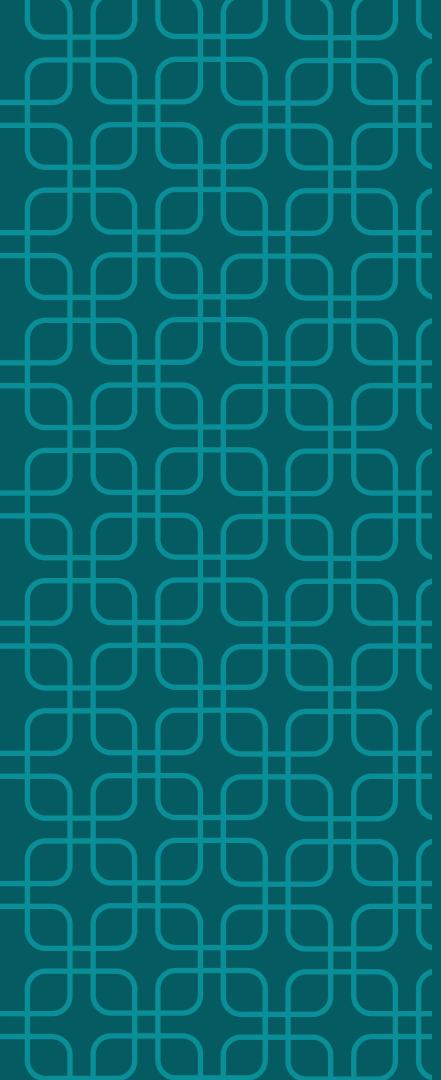
```
[ '<|endoftext|>', 'L', 'aced', 'with', 'dreams', '(', 'dripping', ')', 'in', 'reality', ',',  
'the', 'American', 'Dream', 'reign', 'ites', 'after', '9', '.', '11', 'with', 'a',  
'true', 'story', 'about', 'the', 'Devil', 'Ray', "s", 'mid', '-', 'life', 'rookie', ',',  
'Jimmy', 'Morris', '.', '<|endoftext|>']
```

```
# xlnet-base-cased
```

```
[ '<cls>', '_Lac', 'ed', '_with', '_dreams', '_', '(', '_dripping', ')', '_in', '_reality', ',', '_the',  
'_American', '_Dream', '_reign', 'ites', '_after', '_9', '.', '11', '_with', '_a', '_true',  
'_story', '_about', '_the', '_Devil', '_Ray', "!", 's', '_mid', '(', 'life', '_rookie', ',',  
'_Jimmy', '_Morris', '.', '</s>']
```

Sorting & Batch Training





Q&A

Thank you Covera Health!

Natural Language Interface for Electronic Health Records

Overview - NLI for EHR

1 Introduction

- (1) Motivation of Work
- (2) Data Sources

2 Methods

- (1) Corpus Creation
- (2) Question-SQL Pair Creation
- (3) Named Entity Recognition
- (4) Semantic Parsing

3 Results

- (1) NER on i2b2 Dataset
- (2) NER on Question-SQL Corpus
- (3) Seq2Seq on Question-SQL Corpus

4 Conclusion

- (1) Final Takeaways

Introduction

Motivation of Work: Why a Natural Language Interface for EHR?

- Build a prototype interface on the Medical Information Mart for Intensive Care III (MIMIC III) database, a publicly-available database of EHRs of critical care patients.
- “Natural Language Question” to SQL

Data Sources

- **MIMIC-III Critical Care Database**
 - ~60,000 ICU patients
 - 28 tables: includes admission dates and times, laboratory tests, medications, transfers, and more
- **2012 i2b2 Clinical NLP benchmark dataset**
 - Entity tags include:
 1. problems (“upper quadrant pain”, “diabetes”)
 2. treatments (“blood transfusion”, “aspirin”)
 3. tests (“EKG”, “INR”, “cardiac enzymes”)
 4. clinical departments (“surgery”, “ICU”)
 5. evidentials (“presented” in “Patient presented with...”)
 6. occurrences (events that happen to the patient: “admission”, “transfer”, “follow-up”)

Methods

Question-SQL Pair Creation

Question Types

1. Demographics, which include data such as insurance, race, religion, gender, age
2. Pharmacological Treatments
3. Procedures
4. Comorbidities (*other conditions/diseases patient may have*)
5. Patient Medical History (for example, surgical history)

Question-SQL Pair Dataset Creation

- Crowd-sourced questions
- Constructed “templates”
- Wrote SQL query pairs
- Utilized paraphrasing techniques to extend dataset

How many patients with Diabetes were given Insulin

NER:

How many patients with **PROBLEM@1**
were given **TREATMENT@1**
{PROBLEM@1 : Diabetes,
TREATMENT@1 : Insulin}

Seq2Seq:

```
SELECT count(DISTINCT hadm_id)
FROM DIAGNOSES_ICD
WHERE icd9_code IN
(SELECT DISTINCT icd9_code
FROM D_ICD_DIAGNOSES
WHERE long_title LIKE '%PROBLEM@1%')
AND hadm_id IN
(SELECT DISTINCT hadm_id
FROM PRESCRIPTIONS
WHERE drug LIKE '%TREATMENT@1%')
```

Question-SQL Pair Creation

How many patients suffering from PROBLEM_1 were administered TREATMENT_1

Paraphrases using PPDB ([Paraphrase Database](#)) (Ganitkevitch et al)

How many people suffering from PROBLEM_1 were administered TREATMENT_1

How many patients suffering from PROBLEM_1 were treated TREATMENT_1

Count of patients suffering from PROBLEM_1 were administered TREATMENT_1

```
select count(distinct hadm_id) from DIAGNOSES_ICD where icd9_code IN (select  
distinct icd9_code from D_ICD_DIAGNOSES where long_title like '%PROBLEM_1%') and  
hadm_id IN (select distinct hadm_id from PRESCRIPTIONS where drug like  
'%TREATMENT_1%')
```

Overall Pipeline

Name Entity Recognition

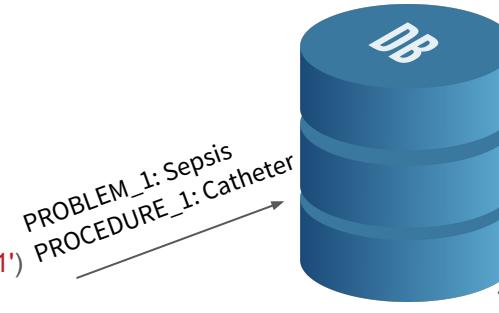
How many patients died with
<PROBLEM_1> after receiving
<PROCEDURE_1>

PROBLEM_1: Sepsis
PROCEDURE_1: Catheter

Seq2Seq with attention

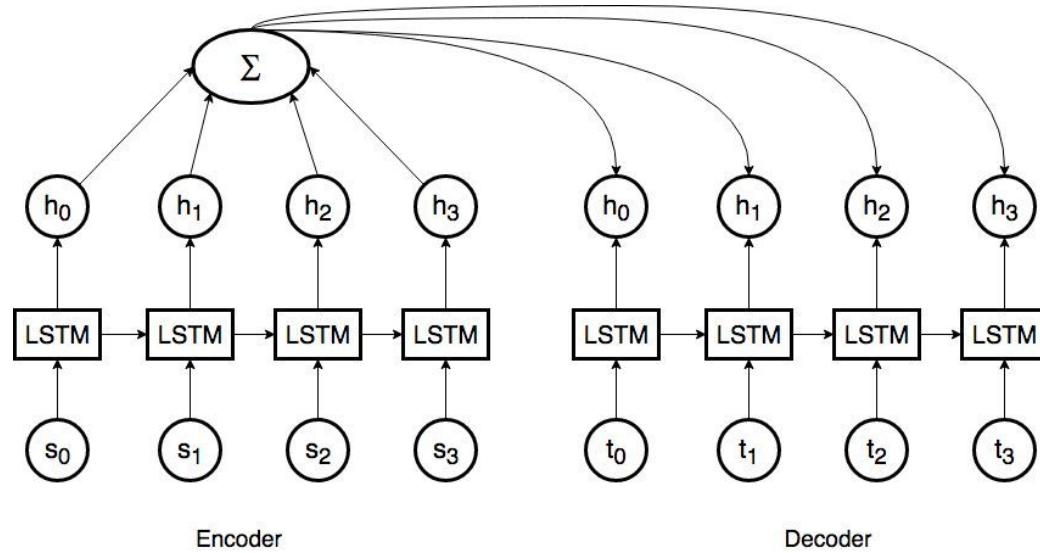
```
select count(distinct subject_id)
from PATIENTS
where expire_flag=1 and subject_id IN
  (select distinct subject_id
   from DIAGNOSES_ICD
   where icd9_code IN
     (select distinct icd9_code
      from D_ICD_DIAGNOSES
      where long_title like 'PROBLEM_1')
   and subject_id IN
     (select distinct subject_id
      from PROCEDURES_ICD
      where icd9_code IN
        (select distinct icd9_code
          from D_ICD_PROCEDURES
          where long_title like 'PROCEDURE_1')))
```

Connect to DB and evaluate



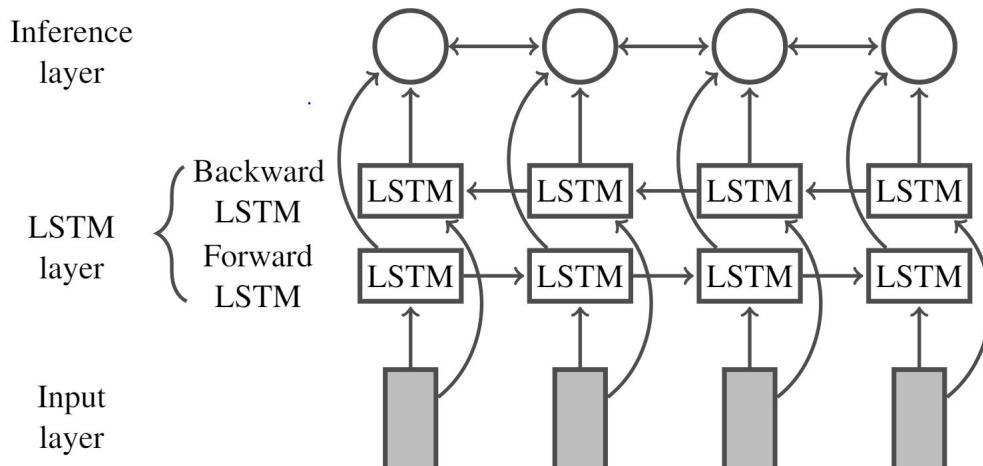
Seq2Seq with attention

- Dong and Lapata(2016) proposed Seq2Seq models with attention for semantic parsing
- Comparatively similar results on GEO(~500) and JOBS(~600)
- Question fed to Encoder. Hidden representation with attention to Decoder



Named Entity Recognition

- Trained on i2b2 2012 clinical NER task (Sun et. al, 2013)
- Dataset has 6 entity types: *problems, tests, treatments, clinical departments, evidentials, occurrences*
- Pretrained word embeddings on MEDLINE and Wikipedia
- Bi-LSTM with Conditional Random Field



Example of NER Output

```
text="medical center" 0:3 --> tag="clinical_dept"
text="physical trauma activation" 9:11 --> tag="problem"
text="w pmhx afib" 24:26 --> tag="problem"
text="coumadin" 28:28 --> tag="treatment"
text="dementia" 30:30 --> tag="treatment" #####
text="breast ca biba s / p fall" 32:38 --> tag="problem"
text="this visit" 79:80 --> tag="problem" #####
text="atrial fibrillation" 141:142 --> tag="problem"
text="guilty about drinking" 295:297 --> tag="problem"
text="bun" 544:544 --> tag="test"
text="creatininine" 546:546 --> tag="test"
text="wbc" 567:567 --> tag="test"
text="hct" 569:569 --> tag="test"
text="plt" 571:571 --> tag="test"
text="results" 579:579 --> tag="occurrence"
text="inr" 590:590 --> tag="test"
text="npo" 732:732 --> tag="treatment"
text="ivf" 734:734 --> tag="treatment"
```

Seq2Seq with Attention Architecture

- Dropout on non-recurrent connections for regularization, as suggested by Pham et al. (2014).
- Beam search is used for decoding the SQL queries after learning

Evaluation

We evaluated our “natural language question” to SQL query in two ways:

1. If NER was correct compared to the golden standard
2. If the SQL executed correctly

Results

Bi-LSTM CRF on i2b2 2012 clinical NER task

- Accuracy = 0.96
- F-1 Score = 0.80

Bi-LSTM CRF on question-SQL corpus

- Accuracy = 0.78
- F-1 Score = 0.75

Seq2Seq with attention on question-SQL corpus

- Accuracy of obtaining SQL = 0.86
- Accuracy of SQL query results = 0.72

Conclusion

Conclusion & Limitations

We built the largest EHR Question-SQL query dataset for MIMIC III as of date.

Most of the question-SQL pairs in this demonstration dataset included fixed-cardinality and simple SQL queries, for example, such as:

“What was the main admission reason for patients who were given lorazepam intramuscularly died?”.

Question-SQL pair dataset includes many variable-length cardinality questions with complex SQL queries for example the pair that matches to the question

“*What patients had a 30% increase in Scr and no dose adjustment to medications?*”.

References

- Alan R Aronson and François-Michel Lang. 2010. An overview of metamap: historical perspective and recent advances. *Journal of the American Medical Informatics Association* 17(3):229–236.
- Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. Mimic-iii, a freely accessible critical care database. *Scientific data* 3:160035.
- Brian L Cairns, Rodney D Nielsen, James J Masanz, James H Martin, Martha S Palmer, Wayne H Ward, and Guergana K Savova. 2011. The mipacq clinical question answering system. In *AMIA Annual Symposium Proceedings*. American Medical Informatics Association, volume 2011, page 171.
- Li Dong and Mirella Lapata. 2016. Language to logical form with neural attention. CoRR abs/1601.01280. <http://arxiv.org/abs/1601.01280>.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* Ellen M Voorhees and William R Hersh. 2012. Overview of the TREC 2012 Medical Records Track. In *TREC*.
- Guergana K Savova, James J Masanz, Philip V Ogren, Jiaping Zheng, Sunghwan Sohn, Karin C Kipper-Schuler, and Christopher G Chute. 2010. Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association* 17(5):507–513.
- J Van Der Volgen, BR Harris, and D Demner-Fushman. 2013. Analysis of consumer health questions for development of question–answering technology. In *Proceedings of one HEALTH: information in an interdependent world, the 2013 annual meeting and exhibition of the Medical Library Association (MLA)*.
- Kirk Roberts, Halil Kilicoglu, Marcelo Fiszman, and Dina Demner-Fushman. 2014a. Decomposing consumer health questions. *Proceedings of BioNLP 2014* pages 29–37.
- Kirk Roberts, Kate Masterton, Marcelo Fiszman, Halil Kilicoglu, and Dina Demner-Fushman. 2014c. Annotating question types for consumer health questions. In *Proceedings of the Fourth LREC Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing*.
- Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention- based neural machine translation. *arXiv preprint arXiv:1508.04025* . Sofia J Athenikos and Hyoil Han. 2010. Biomedical question answering: A survey. *Computer methods and programs in biomedicine* 99(1):1–24.
- Yoav Artzi and Luke Zettlemoyer. 2013. Weakly supervised learning of semantic parsers for mapping instructions to actions. *Transactions of the Association of Computational Linguistics* 1:49–62.

Abbreviations

- MIMIC-III: Medical Information Mart for Intensive Care III
- ICU: Intensive Care Unit
- ICD-9: International Classification of Diseases
- i2b2: Informatics for Integrating Biology & the Bedside
- MeSH: Medical Subject Headlines

Back Up if Interested
in SMM4H Task

SMM4H Shared Task 2020

A Hybrid Pipeline for Identifying Prescription Drug Abuse from Twitter: Machine Learning, Deep Learning, and Post-Processing

Shared Task 4 Overview

- Multi-Class text categorization of tweets mentioning prescription medications as being indicative of potential abuse/misuse (**A**), consumption/non-abuse (**C**), mention-only (**M**), or an unrelated reference (**U**).
- Inter-Annotator Agreement of class labels: 0.681¹
- Evaluation Metric: F1 score for class (A)

	Abuse	Consumption	Mention	Unrelated	Total
Train	1685 (16%)	2940 (28%)	5488 (52%)	424 (4%)	10537
Validation	448 (17%)	730 (28%)	1353 (51%)	104 (4%)	2635

¹ Cohen Kappa between expert annotators (AN1 + AN2): <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7066507/>

Approach Overview

Data Collection & Processing

1

- (1) Additional unlabeled text collection
- (2) Text Preprocessing

Data Augmentation

2

- (1) Snorkel Transformation Functions

Baseline Model

3

- (1) SVM + Feature Engineering

Text Classifier

4

- (1) Word Embeddings
- (2) Pretraining
- (3) Embedding Matrix
- (4) CNN with Attention
- (5) Stacked Logistic Regression

Post-Processing

5

- (1) Exploring Weak Supervision
- (2) Heuristic rules for label corrections

Results & Discussion

6

- (1) Task Performance
- (2) Error Analysis
- (3) Next Steps

Data Collection & Processing

Data Collection

Source	Number of Tweets/Sentences
Twitter Stream Archive (archive.org)	3,106,783
Tweets pulled using original drug terms listed in (O'Connor et al., 2020)	85,188
UCI Drug Review datasets (Gräßer et al., 2018)	1,313,299
Consumer Health Question Answer (CHQA) Corpus (Kilicoglu et al., 2018)	2,595
First 1 billion bytes of English Wikipedia ²	5,244,360

Table 1: Corpora - Text Sources for Unsupervised fastText Word-Embeddings Models

Unlabeled Corpora Expansion

- UCI Drug Datasets
- Consumer Health Question Answer Corpora
- Health Tweets Archive
- Additional pulled tweets using Twitter API
- English Wikipedia

SUBJECT:
EMAIL: [EMAIL]
NAME: [NAME]
GROUP: General Public
STATE: NY
COUNTRY: USA
FROM: /medlineplus/druginfo/natural/786.html
BROWSER: Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.102 Safari/537.36
DATE: 01/31/2014

MESSAGE: I took SAM-E for several months at 400 mg divided 2 a day. It wiped out my depression without the usual numbing side-effect. My mood and intellect were clear after a week. Anxiety disappeared. My vitality, interest and satisfaction returned after years. One night of insomnia in which I was forced to view my age and real failures due to life-long treated depression, lack of a support system (love and family), financial problems,SUDDENLY it stopped working. Went back on it after a month, to no effect? (1)can this be explained?(2) should I go back on it again as nothing else works? (I have been tried on the latest antidepressants by world-famous psychiatrists at the best institution. (I will not identify it for ethical reasons).

Zoloft review by 38 year old female patient

Rating
Overall rating: ★★★★★★★★
Effectiveness: Considerably Effective
Side effects: Mild Side Effects

Treatment Info
Condition / reason: depression;
Dosage & duration: 100 mg. taken every day for the period of 2 years
Other conditions: anxiety;obsessive compulsive thinking
Other drugs taken: none

Reported Results
Benefits: The benefits included mood elevation,obsessive thinking became less. Anxiety and sleeping definately improved.
Side effects: Major weight gain!!!!
Comments: I had to take 25mg. to start and increase from there up to 100 mg. per day. I also had some therapy. Homework. I had to slowly put myself into situations that previously made me very anxious, this drug truly helped!!! I really did hate the weight gain though.

Text Preprocessing

- Tweets were pre-processed using Ekphrasis and annotated for Tokenization

- Word normalization for :
 - Url
 - Email
- Annotation around terms :
 - hashtags
 - allcaps
 - elongated
 - repeated
 - emphasis
 - censored
- User
- Percent
- Time and date terms

Examples:

Original Raw Tweet Text	Preprocessed Text	Label
1.2 mg .02 of my x 6 i feeeeeeeeel goooooood like im on suboxone or methadone	<number> mg . <number> of my x <number> i feel <elongated> good <elongated> like im on suboxone or methadone	ABUSE
Should I Up my klonopin dose my prescription is .5-1 twice a day as needed and I only do .5 usually should I do 1	should i up my klonopin dose my prescription is . <number> - <number> twice a day as needed and i only do . <number> usually should i do <number>	CONSUMPTION
Alexander Jean-Whiskey and Morphine #iHeartAwards #BTSARMY #BestFanArmy _U	alexander jean - whiskey and morphine <hashtag> i heart awards </hashtag> <hashtag> btsarmy </hashtag> <hashtag> best fan army </hashtag> <user>	UNRELATED
Heroin, morphine #iHeartAwards #BestSoloBreakout #CamilaCabello	heroin , morphine <hashtag> i heart awards </hashtag> <hashtag> best solo breakout </hashtag> <hashtag> camila cabello </hashtag>	MENTION
_U xanax please. i'll still laugh...but i'll totally take the xanax...	<user> xanax please . i will still laugh . <repeated> but i will totally take the xanax . <repeated>	ABUSE

Data Augmentation

Data Augmentation

Transformation functions

Developed custom synonym dictionaries to avoid altering the label of the tweet

- Slang terms were replaced only by other slang terms
 - Avoided ambiguous slang terms, e.g. ‘bananas’ is used as a slang term for oxycodone,¹ but swapping with ‘oxy’ could alter tweet label
“**Bananas** are my favorite fruit” → “**Oxy** are my favorite fruit”
- Drug names replaced with other drug names
 - Swapped only with others in the same drug class
 - Exception: select slang terms swapped with drug terms
 - Some slang terms can be used in place of a medication name without changing the meaning of the tweet, e.g ‘benzo’ is not a proper term, but can semantically be interchanged with a benzodiazepine drug name (lorazepam, alprazolam, etc)
- Verbs replaced by a predefined set of synonyms
 - Verbs associated with abuse only swapped with others e.g. “crush” → “snort”

¹ 2018 Slang Terms and Code Words - DEA : <https://www.dea.gov/sites/default/files/2018-07/DIR-022-18.pdf>

Data Augmentation

Replace words with synonyms

winnipeg rules to me, personally, because i am a **big** fan of murder and oxycontin abuse.
winnipeg rules to me, personally, because i am a **large** fan of murder and oxycontin abuse.

Replace references to relatives

my **brother** knows all the kiss me kiss me lyrics bc i listen to it so much
my **dad** knows all the kiss me kiss me lyrics bc i listen to it so much

Replace verbs indicative of abuse

Ay mane how I dont know u connected, heres a little morphine let me see u **inject** it
Ay mane how I dont know u connected, heres a little morphine let me see u **snort** it

Replace drugs of similar class

ali as after his first **xanax**
ali as after his first **alprazolam**

Baseline Model

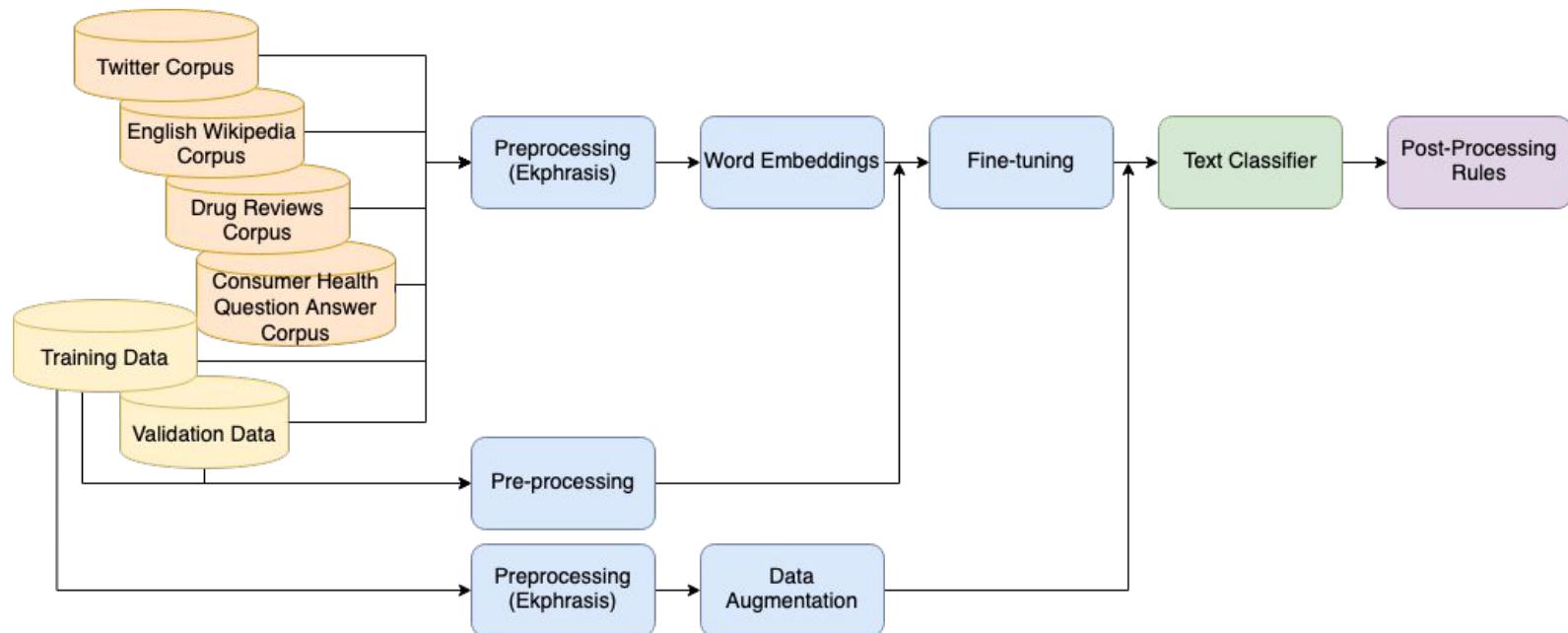
Support Vector Classifier

- Linear Kernel
- One-vs-Rest Decision Function
- Features (*extracted after pre-processing with ekphrasis*)
 - Vader Sentiment Score
 - Binary flag features including presence of commonly abused drug terms, and presence of emojis commonly associated with drug abuse
 - Count-based features such as number of hashtags, username mentions, emojis/emoticons and number of drug terms identified
 - Co-occurrence of chemical and disease entities, where entity extraction was performed using bc5cdr named entity recognizer
 - Term Frequency-Inverse Document Frequency (tf-idf) of tokens (unigrams and bigrams)

	F1	Precision	Recall
ABUSE	0.49	0.5276	0.4553

Deep Learning CNN Text Classifier

System Pipeline



CNN Based Text Classifier

- Pre-trained Word Representations
- Pre-training with labeled text (validation + train combined)
- Embedding Layer
- 2 layer CNN with Max Pooling and Parametric Attention
- Stacked Unigram Logistic Regression

Embedding Layer

- Pre-training (params: conv depth of 4, cosine loss function, drop out of 0.4, and batch size of 64) on the training dataset with the word embeddings to create weights that initialize our embedding matrix
- Additionally our embedding layer also utilizes linguistic features such as part-of-speech, dependency, word-shape, etc.

token.text	token.lemma_	token.pos_	token.tag_	token.dep_	token.shape_	token.is_alpha	token.is_stop
1.2	1.2	NUM	CD	nummod	d.d	False	False
mg	mg	PROPN	NNP	ROOT	xx	True	False
.02	.02	NUM	CD	appos	.dd	False	False
of	of	ADP	IN	prep	xx	True	True
my	-PRON-	DET	PRP\$	poss	xx	True	True
x	x	SYM	SYM	punct	x	True	False
6	6	NUM	CD	ROOT	d	False	False
i	i	PRON	PRP	nsubj	x	True	True
eeeeeeeeeeel	eeeeeeeeeeel	VERB	VBP	ROOT	xxxx	True	False
oooooooood	oooooooood	NOUN	NN	acomp	xxxx	True	False

Post-Processing

Generative Labeling Functions

Initial Goal: Enrich training data with supplemental tweet corpus

- Use Weak Supervision for labelling new data
 - Utilize SME knowledge to develop set of heuristic rules
 - Perform noise-aware labeling using codified rules + Snorkel
- Result: Low coverage of rules led to low label accuracy/coverage

Updated Goal: Use subset of heuristic rules for post-processing

- Use rules to override class prediction of model
 - Keep only rules with acceptably high empirical accuracy (>65%)
 - Order by relative precision and have rules either override tweet label or pass to next rule.

Defining Heuristics

Initial Generation of rules

Examples

Abuse

Pleasantly surprised when finding a drug
"lol I just found an adderall in my pocket"

Manipulating drug absorption and pharmacokinetics
"Pro tip" on taking something
Mix something with something else to increase its effect
E.g. taking Adderall with an alkaline drink/baking soda,

Mention of a medication with clinically inappropriate routes or indications

Mention

Mention of a med as a means to tell someone to calm down
*"chill out, take a xanax"
"you need a xanax"
"bruh take a xanax, it's twitter'*

Challenges

Indications of consumption/ingestion

- Not all medication use is associated with a verb.
- Defined other drug/word associations - mention of a time of past/future ingestion or specific meal:
 - E.g. [number]PM/AM, for breakfast, at bedtime, before bed
*"valium for breakfast, kl"
"it's time for my 5pm xanax"*

Appropriate vs. inappropriate usage

- Varies by drug and approved indication
- Can also depend on if tweet clearly indicates usage is as-prescribed even if off-label
"told my dr i'm having trouble sleeping. he gives me xanax and diazepam. i'm bout to be a walking zombie...."

Terminology dictionaries

- Building comprehensive list of drug names/aliases, routes, conditions, verbs for consumption
- Categorizing verbs, (drug, conditions) pairs by suspected abuse / non-abuse

Development Challenges

Example of rule with low accuracy due to broad voting criteria

- Use_with_caffeine (empirical accuracy: 0.55)
- Error analysis
 - Stimulant + coffee/caffeine/red bull → abuse
 - Spiked, put(s) in, slip(ped) in, added + DRUG → mention
 - Benzo + coffee/caffeine → consumption
 - Unless another drug or alcohol is mentioned→ abuse

difficulty differentiating:
co-occurrence vs actual co-ingestion
actual co-ingestion (or intent) versus
figurative statements

Text	Groundtruth	Vote	Correct?
eyes are rolling think i'll get coffee before get my lorazepam blues !!!!	CONSUMPTION	ABUSE	FALSE
it's funny how my allowed one cup of coffee gets me going like vyvanse gets everyone else going..... srsly	MENTION	ABUSE	FALSE
coffee and vyvanse for a pick me up.	ABUSE	ABUSE	TRUE
Coffee & Klonopin is what my mornings are made of.	CONSUMPTION *	ABUSE	FALSE
did someone slip an adderall in my coffee or what	MENTION*	ABUSE	FALSE

* Based on the task annotation guideline, would expect the groundtruth label to be 'ABUSE'

Labeling Functions

Programming Heuristic Rules

LOGIC:

```
- if is the word lyrica mentioned in tweet
  - if the number of unique possible drug mentions is 1 (e.g.lyrica)
    - check that we find at least one term in the tweet associated with the artist
    - check that we do not find a regex pattern that matches dose information for a drug
    - if we meet the above 2 criteria
      - override P and return UNRELATED
    else
      - return P
  else
    - if P is UNRELATED (even though we see another likely drug term)
      - override P and return MENTION
    else
      - return P
else
  - return P
```



```
def lyrica_anderson(doc, P, score=None, threshold=1, pass_p=True):
    '''search for terms that are likely to be associated with Lyrica the artist'''
    PASS = P if pass_p else None

    person_terms = {'anderson', 'lhhh', 'lgbt', 'lgbtq',
                    'song', 'album', 'sing', 'sang', 'singing', 'record',
                    'hip hop', 'hiphop', 'teairra', 'moniece', 'masika',
                    'safaree', 'omarion'}
    drug_terms = {'prescription(s)?', 'fybromyalgia', 'diabet(es|ic)',
                  'shingles', 'medicat(ion(s)?|ed)', 'seizure(s)?',
                  'pharmac(y|ist)*',
                  r'\d+[ ]?(mg|mcg|ml|mcL|pills|tablet(s)?)'}
    drug_rex = ('|' + '|'.join(drug_terms) + ')'

    # pass if we dont see lyrica in the tweet
    found_drugs = [ent.text for ent in doc.ents if ent.label_=='DRUG']
    if 'lyrica' not in found_drugs:
        return PASS
    # if lyrica in the tweet, but there's another possible drug, we would rather pass
    elif len(set(found_drugs))>1:
        # if the model predicts unrelated we may want to override with some more likely class,
        # like MENTION since a second possible drug is not likely to be referring to a person
        if P==UNRELATED:
            return MENTION
        else:
            return PASS
    # check if we see any key terms we know are associated with the person
    # also check that the text does not match a typical dose string pattern.
    # if we pass both criteria we will override and return UNRELATED, otherwise pass the original pred
    tokens = set([t.text for t in doc])
    if len(tokens.intersection(person_terms))>0 and not re.search(drug_rex, doc.text):
        return UNRELATED
    else:
        return PASS
```

Post-Processing

	j	Polarity	Coverage	Overlaps	Conflicts	Correct	Incorrect	Emp. Acc.
drug_with_slang_usage	0	[0]	0.019645	0.019645	0.019645	87	120	0.420290
drug_with_consumption_usage	1	[0]	0.100978	0.100978	0.100978	210	854	0.197368
no_drugnames_found	2	[3]	0.004176	0.004176	0.004176	6	38	0.136364
...								
drug_had_me	30	[2]	0.011958	0.011958	0.011958	87	39	0.690476
died_of	31	[1]	0.001139	0.001139	0.001139	10	2	0.833333
lil_xan	32	[1, 3]	0.002373	0.002373	0.002373	21	4	0.840000
addicted_3rdperson	33	[0, 1, 2]	0.031603	0.031603	0.031603	212	121	0.636637
clinical_terms	10	[2]	0.052861	0.052861	0.052861	253	304	0.454219
medical_mention_terms	11	[1]	0.023441	0.023441	0.023441	157	90	0.635628
benzo_specific_phrases	12	[0]	0.000759	0.000759	0.000759	5	3	0.625000
abusive_activities	13	[0]	0.023441	0.023441	0.023441	88	159	0.356275
study_aid	14	[]	0.000000	0.000000	0.000000	0	0	0.000000
soliciting_drug	15	[0]	0.031128	0.031128	0.031128	95	233	0.289634
suspicious_emojis	16	[]	0.000000	0.000000	0.000000	0	0	0.000000
prescribed_indication	17	[]	0.000000	0.000000	0.000000	0	0	0.000000
expected_usage	18	[]	0.000000	0.000000	0.000000	0	0	0.000000
lyrica_as_nondrug	19	[3]	0.020404	0.020404	0.020404	190	25	0.883721

- Label Functions (LF) with **>65%** accuracy were selected
- Selected LFs were ordered by decreasing precision
- The LFs are applied one-by-one to the a tweet after it has passed through the model
- Each LF can either override the class label predicted by the model, or pass to the next LF in line
- The final class label is set by the first LF to override, or the model's original predicted label is used if all LFs pass

Results & Discussion

Model Performance & Error Analysis

CNN Deep Learning Model
performance on validation set :

	precision	recall	f1-score	support
ABUSE	0.732	0.634	0.679	488
CONSUMPTION	0.781	0.858	0.817	730
MENTION	0.882	0.876	0.879	1353
UNRELATED	0.931	0.913	0.922	104

- Our system performed an f1 score of 0.51- recall an upper-bound = **0.681¹**
- Abuse tweets were commonly labeled as mention or consumption
 - The model struggled to classify the abuse of a relative or close friend, perhaps because these types of tweets may have been lacking in the training data
- Among mislabeled abuse tweets, the most common medication class was benzodiazepines, with Xanax being the most commonly mislabeled drug
 - Xanax is often used in everyday vernacular without being indicative of drug abuse (i.e. as a figure of speech expressing the need to calm down), making it more challenging for the model to delineate subtle differences in tweets' semantic meaning

Performance on **official test data** (*ABUSE* class):

	precision	recall	f1-score
Baseline <i>SVC + Feature engineering</i>	0.5276	0.4553	0.49
Text Categorizer <i>CNN w/ Attention + stacked Linear + post-processing</i>	0.5306	0.4831	0.51

■ *"If one more entitled Xanax mom yells at me to honor a coupon that expired 3 years ago I'm jumping off a cliff"*

Next Steps

- Enrichment with Outside Knowledge Sources
 - UMLS, MedDRA, SIDER, HLP Lab ADR Lexicon, etc.
 - Incorporating outside knowledge, for example, lyrics, pop culture, etc(reference glue tasks for common sense knowledge & world knowledge)
- Labelling Function Development
 - Continued work with SMEs to refine heuristic rules and improve coverage of dataset
 - Weak Supervision (Snorkel) for labelling supplemental downloaded tweets
- Data Evaluation
 - Evaluate dataset and word embeddings for bias
 - Confidence levels for ground truth abuse/misuse labels to differentiate between high confidence of abuse/misuse versus subjective hints of abuse
 - Descriptive analysis to further categorize tweets, types of abuse, to identify emerging themes, for example, “informational content”, “seeking treatment”, “sarcasm/jokes”, etc.

References

F. Gräßer, S. Kallumadi, H. Malberg, and S. Zaunseder, "Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning," in *Proceedings of the 2018 International Conference on Digital Health*, 2018, pp. 121–125, doi: 10.1145/3194658.3194677.

J. W. Wei and K. Zou, "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks," *CoRR*, vol. abs/1901.1, 2019, [Online]. Available: <http://arxiv.org/abs/1901.11196>.

P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *arXiv Prepr. arXiv1607.04606*, 2016.

H. Kilicoglu *et al.*, "Semantic annotation of consumer health questions," *BMC Bioinformatics*, vol. 19, no. 1, 2018, doi: 10.1186/s12859-018-2045-1.

A. Sarker, G. Gonzalez-Hernandez, Y. Ruan, and J. Perrone, "Machine Learning and Natural Language Processing for Geolocation-Centric Monitoring and Characterization of Opioid-Related Social Media Chatter," *JAMA Netw. Open*, vol. 2, no. 11, 2019, doi: 10.1001/jamanetworkopen.2019.14672.

H. Hu *et al.*, "An insight analysis and detection of drug-abuse risk behavior on Twitter with self-taught deep learning," *Comput. Soc. Networks*, vol. 6, no. 1, p. 10, Nov. 2019.

Drugs.com, "Phonetic and Wildcard Search at Drugs.com." 2020.

Substance Abuse and Mental Health Services Administration, "Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health (HHS Publication No. PEP19-5068, NSDUH Series H-54)," Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Rockville, MD, 2019.

T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality." 2013.

X. Yang, C. Macdonald, and I. Ounis, "Using Word Embeddings in Twitter Election Classification." 2016.

Y. Kim, "Convolutional Neural Networks for Sentence Classification." 2014.

References

- U.S. Food and Drug Administration, "Pharmacologic Class," *U.S. Food and Drug Administration*. 2018.
- S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning Based Text Classification: A Comprehensive Review." 2020.
- A. Sarker, A. DeRoos, and J. Perrone, "Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework," *J. Am. Med. Inform. Assoc.*, vol. 27, no. 2, pp. 315–329, Feb. 2020.
- J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," *Int. J. Res. Mark.*, vol. 36, pp. 20–38, 2019.
- A. Z. Klein *et al.*, "Overview of the fifth Social Media Mining for Health Applications (#SMM4H) Shared Tasks at COLING 2020," 2020.
- Kowsari, J. Meimandi, Heidarysafa, Mendum, Barnes, and Brown, "Text Classification Algorithms: A Survey," *Information*, vol. 10, no. 4, p. 150, Apr. 2019, doi: 10.3390/info10040150.
- C. Baziotis, N. Pelekis, and C. Doulkeridis, "DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Aug. 2017, pp. 747–754, doi: 10.18653/v1/S17-2126.
- W. Yin, H. Schütze, B. Xiang, and B. Zhou, "ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs," *CoRR*, vol. abs/1512.0, 2015, [Online]. Available: <http://arxiv.org/abs/1512.05193>.
- Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical Attention Networks for Document Classification," 2016.
- A. Ratner, B. Hancock, J. Dunnmon, F. Sala, S. Pandey, and C. Ré, "Training Complex Models with Multi-Task Weak Supervision." 2018.
- L. Zhining, G. Xiaozhuo, Z. Quan, and X. Taizhong, "Combining Statistics-Based and CNN-Based Information for Sentence Classification," *2016 IEEE 28th Int. Conf. Tools with Artif. Intell.*, pp. 1012–1018, 2016.
- Y. Zhang and B. Wallace, "A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification." 2016.

Methods

Why we are generating artificial chemicals (the drug space of toxicity)

Chemical/Drug structure dictates how a drug works --> toxicity can be determined via the drug's structure. We measure toxicity by LD50, the median lethal dose of the drug

Conventionally, QSARs (Quantitative Structure Activity Relationships) are used and follow the form of a linear combination of molecular descriptors. These can be difficult to use.

For all architectures

- All architectures implemented in PyTorch
- Adam Optimizer
- Binary Cross Entropy Loss Function
- Early Stopping
- Max Epochs: 200
- Vocabulary Size: 20,000
- Input Length: 2,000

CNN architectures - hyperparameters

Hyperparameters tested for CNN architectures

- Embedding Dimension: {100, **300**}
- Number of Kernels: {50, 100, **200**}
- Kernel Size: {**5**, 10}
- Activation Function: {**relu**, tanh}
- Learning rate: {1e-3, **3e-3**}

Grid search optimization used

CNN architectures (cont'd)

- Max Pooling
- One Fully Connected Layer
- Mini-Batch Size: 512
- Dropout: 0.2

Grid search optimization used

RNN architectures

Hyperparameters tested

- Embedding Dimension: {100, **300**}
- Recurrence Unit: {LSTM, **GRU**}
- Hidden Layers: {128, 256}
- Number of Layers: {1, 2}
- Learning rate: {1e-3, **3e-3**}

Grid search optimization used

Activation function: tanh

Text-mining for automatic contract

- Well d
- The sdaadaection of the notes omits other information which may be useful for the RNN for example.
- Pre

Limitations & Future Directions

- Well documented issues of mistrust (Boag et al. 2018) and undertreatment (Parikh et al. 2020) by race or ethnicity are likely embedded in H&P notes.
- The segmentation of the History of Present Illness section of the notes omits other information which may be useful for the RNN for example.
- Preprocessing text masks numbers and strips symbols which may carry useful, albeit highly specific, information (Cruz Diaz and Mana Lopez 2015) such as medication or radiation dosage.

Limitations & Future Directions

- Well documented issues of mistrust (Boag et al. 2018) and undertreatment (Parikh et al. 2020) by race or ethnicity are likely embedded in H&P notes.
- The segmentation of the History of Present Illness section of the notes omits other information which may be useful for the RNN for example.
- Preprocessing text masks numbers and strips symbols which may carry useful, albeit highly specific, information (Cruz Diaz and Mana Lopez 2015) such as medication or radiation dosage.

Limitations & Future Directions

- Well documented issues of mistrust (Boag et al. 2018) and undertreatment (Parikh et al. 2020) by race or ethnicity are likely embedded in H&P notes.
- The segmentation of the History of Present Illness section of the notes omits other information which may be useful for the RNN for example.
- Preprocessing text masks numbers and strips symbols which may carry useful, albeit highly specific, information (Cruz Diaz and Mana Lopez 2015) such as medication or radiation dosage.

Limitations & Future Directions

- Well documented issues of mistrust (Boag et al. 2018) and undertreatment (Parikh et al. 2020) by race or ethnicity are likely embedded in H&P notes.
- The segmentation of the History of Present Illness section of the notes omits other information which may be useful for the RNN for example.
- Preprocessing text masks numbers and strips symbols which may carry useful, albeit highly specific, information (Cruz Diaz and Mana Lopez 2015) such as medication or radiation dosage.