Medical Billing Virtual Assistant (MBVA)

The future of medical coding

MIDS Capstone Project Presentation (Aug, 2020)
Brent Biseda | Gaurav Desai | Haifeng Lin | Anish Philip

Berkeley SCHOOL OF INFORMATION

Why Do I Need to Pay for Medical Bill Errors?

80% **bills with error***

$68,000,000,000

Cost of medical bill errors annually*

Source: 4 in 5 Medical Bills Contain Errors: Here's What You Can Do

A Manual Process to Find Medical Billing Codes

Discharge Summary

History of Present Illness (HPI): Attention difficulties she will talk to him face to face and she is unable to pay attention to him. He was diagnosed with <mark>Autism Spectrum Diagnosis</mark> (he was diagnosed with PDD). He is going in the 5th grade. Treatment Options: 10 yo with history of developmental delays, meets criteria for <mark>ADHD and ASD</mark>. He also has anger issues and anxiety, and may meet criteria for OCD. Strong family history of mood and ADHD. Provided <mark>psychoeducation about OCD</mark> and anxiety and ADHD Start Zoloft 25mg. r/b/a including black box warning discussed...
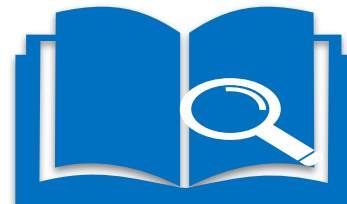
Medical Chapter

ICD 10 Code

What Do Medical Coders Say about Coding Error?

Our Goal

Reduce the error and time taken to code medical diagnosis

MBVA

Our Approach

MBVA

Clinical BERT embeddings

Web App: AWS

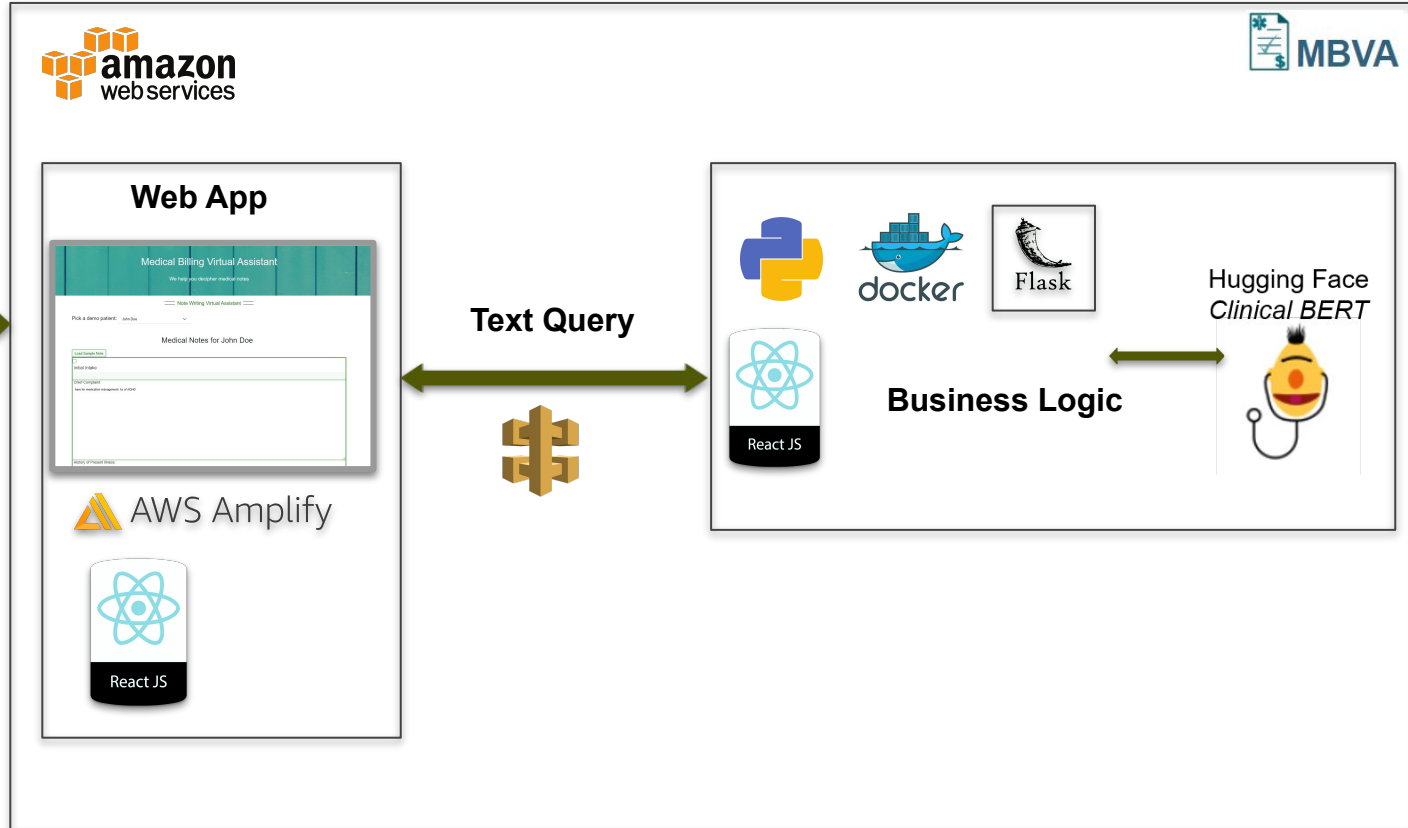**Data** **Preprocess** **Model** **UI**

MIMIC-III

Clinical notes and diagnosis codes linked to each note

CNN / Transfer learning

MBVA Application Architecture

- **Medical Coder**
- **Physician Assistant**
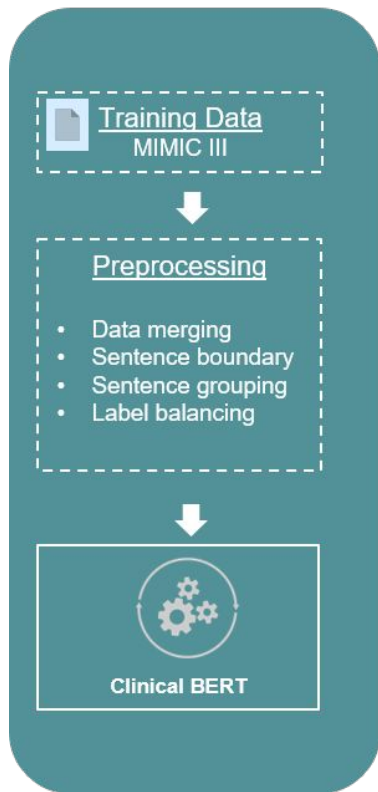- **Nurse Practitioner**
- **Doctor**
- **Patient**

Divide and Conquer with An Ensemble of Multi-Label Classifiers

Preprocessing

Predict ICD Chapters

Predict Disease / Indications

Training Data
MIMIC III

Preprocessing

- Data merging
- Sentence boundary
- Sentence grouping
- Label balancing

Clinical BERT
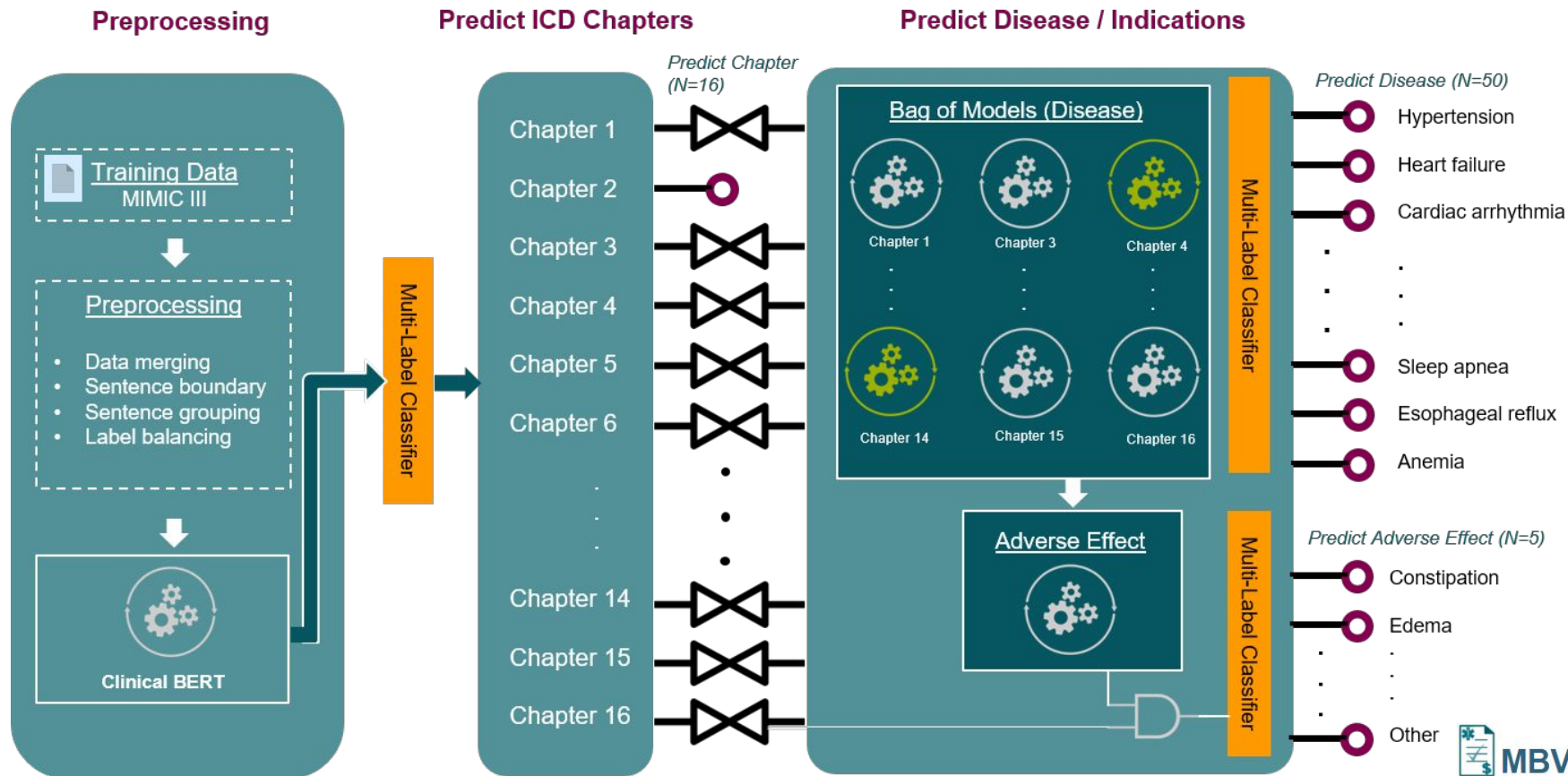
MBVA

Divide and Conquer with An Ensemble of Multi-Label Classifiers
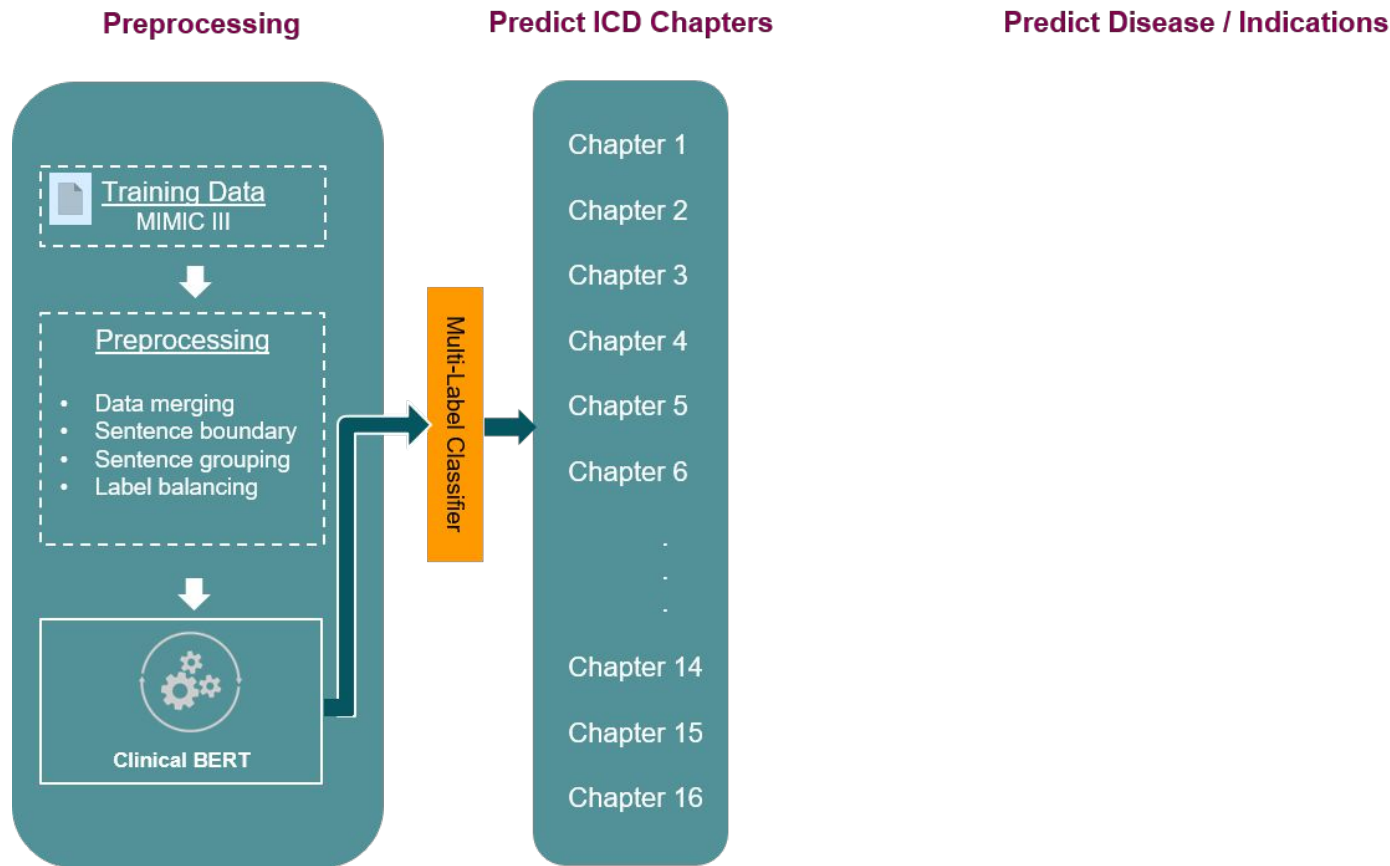
Divide and Conquer with An Ensemble of Multi-Label Classifiers

F1 Scores - Because Both Precision and Recall Are Important

Weighted Avg F1 score over model evolution

Legend:
- Layer 1: Chapters / Therapy areas
- Layer 2: Disease categories
- Single Layer models

Top-100 ICD-9 Codes	16 Chapters + Sliding window classifier	16 Chapters + *Top-68 Disease category + LSTM*	16 Chapters + Top-68 Disease category + *CNN + Transfer learning*	16 Chapters + Top-68 Disease category + CNN + Transfer learning + *Balanced Labels*	16 Chapters + Top-50 Disease category + CNN + Transfer learning + Balanced Labels + *Text Augmentation (4X)*
0.2	0.56	0.62 / 0.2	0.61 / 0.53	0.68 / 0.61	0.78 / 0.75

Final model Tested on 3 batches with 10k unseen documents each

Comparison to Others

Publications / Previous work	Prediction	F1 Scores
Explainable Prediction of Medical Codes from Clinical Text **Association for Computational Linguistics, June 2018**	Secondary prediction: Top 50 ICD codes	**0.63**
Tagging Patient Notes With ICD-9 Codes ***29th Conference on Neural Information Processing Systems***	19 High level ICD-9 Categories	**0.71**
An empirical evaluation of deep learning for ICD-9 code assignment using MIMIC-III clinical notes ***Computer Methods and Programs in Biomedicine, May 2019***	Top-10 ICD-9 codes	**0.72**
Multimodal Machine Learning for Automated ICD Coding ***Proceedings of Machine Learning Research, Aug 2019***	Top-32 ICD-9 codes	**0.76**
Medical Billing Virtual Assistant (MBVA) MBVA ***MIDS Capstone 2020***	● 16 therapy areas ● Most prevalent Disease category across these therapy areas (N=50)	Chapter / Therapy Areas: **0.78** Disease category@50: **0.75**

Medical Billing Virtual Assistant (MBVA)
Preparing to shape the future of medical coding

Sample Note
Neonatal Discharge Summary

Benefits of MBVA

MBVA

HOSPITAL BILLING TEAM
Fewer Error and Higher Productivity

PHYSICIAN/ NURSE
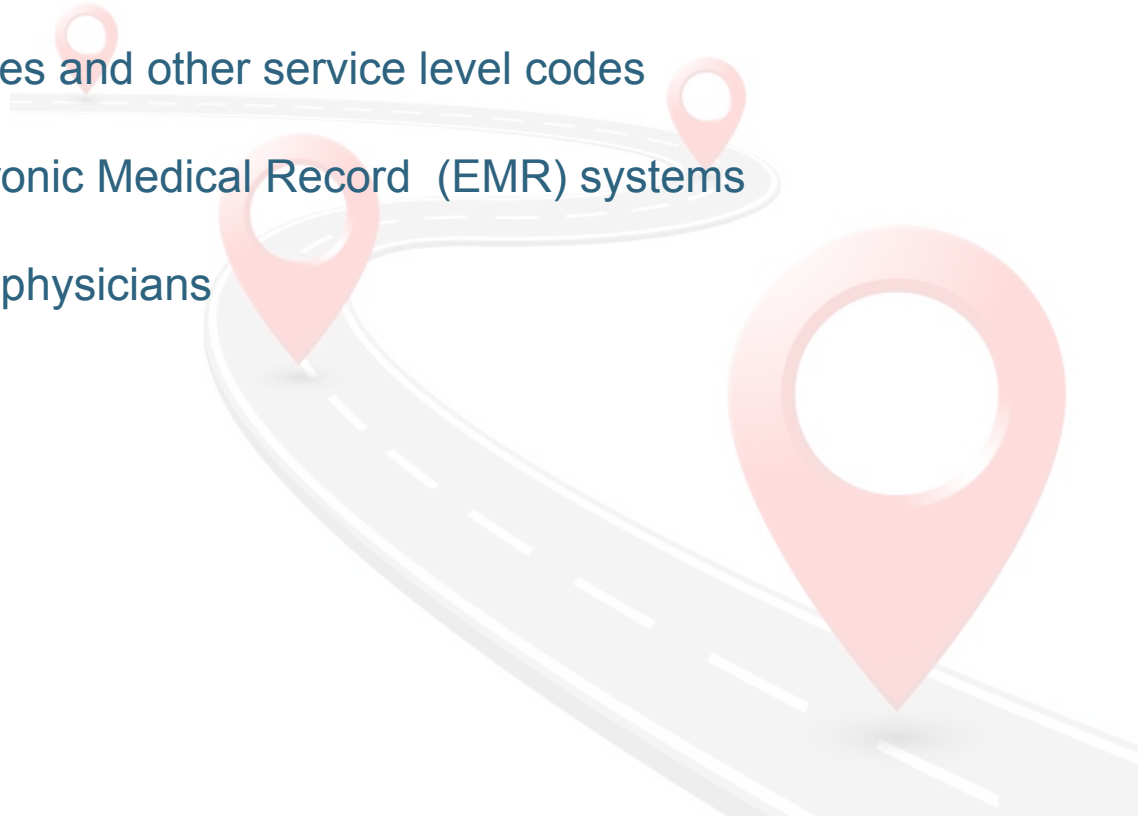More time on patients
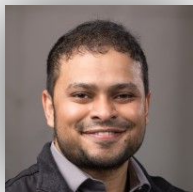
INSURERS
Low cost on Audit

PATIENTS
More transparency on medical charges

Future Roadmap

- Improve model predictions

- Expand to medical procedures and other service level codes

- Integrate with existing Electronic Medical Record (EMR) systems

- Build AI writing assistant for physicians

The Team

Anish Philip
Medical SME and Model Builder
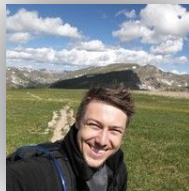
Brent Biseda
Product Manager

Gaurav Desai
UI Developer

Haifeng Lin
Project Manager

Medical Billing Virtual Assistant (MBVA)

Preparing to shape the future of medical coding

Q&A

Thank you

Team Responsibility

Team Member Accomplishment - Brent Biseda

01 Determine initial project scope

02 Obtain SME inputs

03 Decide product design and features

04 Create Flask Backend

Team Member Accomplishment - Gaurav Desai

01 Create UI

02 Decide product design and features

03 Setup AWS environment for the project

Team Member Accomplishment - Anish Philip

01 Conduct EDA and feature engineering

02 Build, run, and decide ML models

03 Our SME in healthcare industry

04 Final presentation PPT (Branding and Structure)

Berkeley
UNIVERSITY OF CALIFORNIA

Team Member Accomplishment - Haifeng Lin

01 Construct ppt for regular update and graded presentation

02 Manage project and coordinate meetings

03 Create contents for the UI

Berkeley
UNIVERSITY OF CALIFORNIA

Appendix

Preprocessing Clinical Text from MIMIC-III

Merging & Preprocessing

Merge multiple notes from the same hospital stay

Preprocess:

- Remove anonymized name/ ID tags
- Replace medical Rx abbreviations with plain English
- Other typical text preprocessing steps

Sentence boundaries and grouping

A note from a single hospital stay

- BERT tokenizer & token to ID (30K vocab)
- N sentence groups with 128 tokens each
- Zero pad in case unable to group adjacent sentences with exactly 128 tokens
- Truncate if a single sentence is >128 tokens

BERT Embeddings

Tokenized Nx128 matrix

Attention Nx128 matrix

Pretrained Clinical BERT >100M trained parameters

128 Tokens

N Sentence grouping

52 k Notes

> 250 GB Vectorized Notes
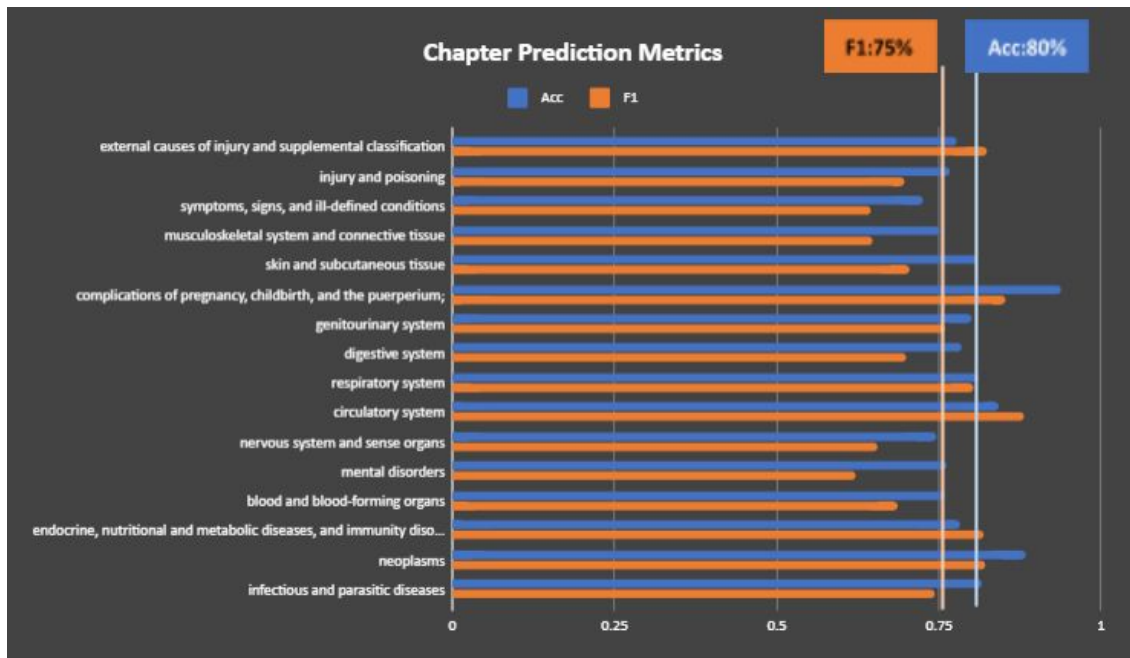
768 floating point representation/ Token

Results and Evaluation

Cut-offs at 0.5 yields the best model performances

Results and Evaluation

Model performances are relatively consistent across 16 chapters, with average F1 score of **75%** and Accuracy of **80%**

Individual Model Performance

Therapy Area/ Chapter	Chapter Description	# of Labels	Training		Test	
			Acc	F1	Acc	F1
001 139	infectious and parasitic diseases	4	88%	80%	78%	62%
140 239	neoplasms	3	85%	82%	78%	73%
240 279	endocrine, nutritional and metabolic diseases, and immunity disorders	4	80%	70%	73%	57%
280 289	blood and blood-forming organs	3	86%	78%	75%	62%
290 319	mental disorders	4	81%	59%	73%	31%
320 389	nervous system and sense organs	6	86%	53%	80%	32%
390 459	circulatory system	6	81%	64%	75%	52%
460 519	respiratory system	5	78%	52%	75%	45%
520 579	digestive system	5	84%	60%	82%	57%
580 629	genitourinary system	3	77%	74%	69%	63%
630 679 740 759 760 779	complications of pregnancy, childbirth, and the puerperium;	4	77%	71%	76%	71%
680 709	skin and subcutaneous tissue	-	-	-	-	-
710 739	musculoskeletal system and connective tissue	2	82%	83%	58%	60%
780 799	symptoms, signs, and ill-defined conditions	5	85%	54%	82%	48%
800 999	injury and poisoning	4	88%	73%	82%	57%
E V	external causes of injury and supplemental classification	7	88%	70%	84%	64%
Adverse Effect		4	74%	45%	72%	37%

Medical Billing Virtual Assistant

We help you decipher medical notes

Paste your medical notes here

Get Codes

Berkeley
UNIVERSITY OF CALIFORNIA

ICD Code Prediction

	ICD 9 Chapter/s	ICD 10 Chapter/s	Description	Probability > 0.35
⌄	Adverse	Adverse	Adverse effect diseases	1
⌃	320-389	H00-H59; G00-G99	Diseases of the nervous system and sense organs	0.55

Diseases

Name	ICD 9 Code	ICD 10 Code	Probability
Paralysis	5601	B0601	0.52
Iritis or other eye disorder	37963	H5943	0.44
Sleep apnea	51881	J9600	0.42
Neuropathy in diabetes	3572	E0842	0.41

	ICD 9 Chapter/s	ICD 10 Chapter/s	Description	Probability
⌄	460-519	J00-J99	Diseases of the respiratory system	0.48
⌄	001-139	A00-B99	Infectious and parasitic diseases	0.46
⌄	290-319	F01-F99	Mental disorders	0.44
⌄	780-799	R00-R99	Symptoms, signs, and ill-defined conditions	0.43
⌄	800-999	S00-T88	Injury and poisoning	0.38
⌄	E-V	S00-T88; V00-Y99; Z00-Z99	External causes of injury and supplemental classification	0.37
⌄	390-459	I00-I99	Diseases of the circulatory system	0.36

Pick a demo patient: [Start typing patient name here ▾]

Medical Notes

[Load Sample Note]

Initial Intake

Chief Complaint:

Chief Complaint

History of Present Illness:

HPI

Berkeley
UNIVERSITY OF CALIFORNIA