NLP Assisted Coding

Amanda Potter

Introduction

Medical coders translate physician notes into a set of diagnosis and procedure codes. These codes are used to send a claim to the payer for reimbursement

Physician Notes

Recorded by physician.

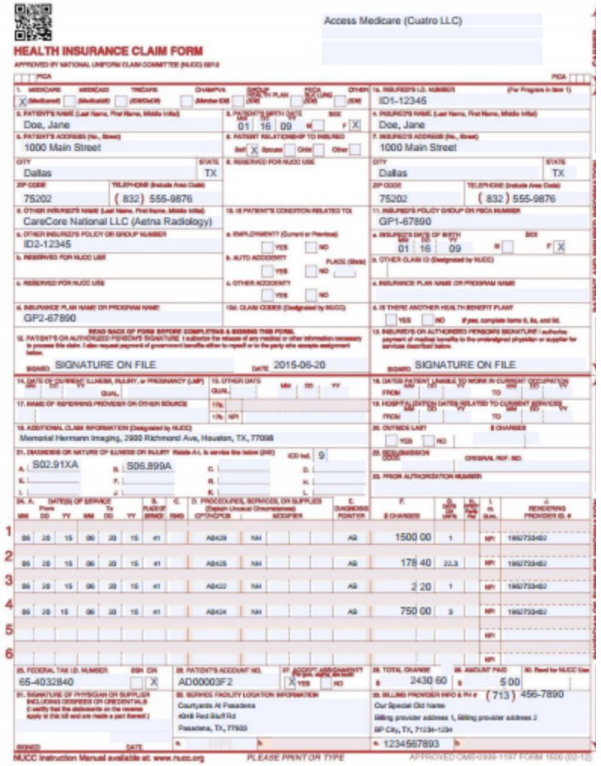
Transcription

Transcriptionist or voice recognition.

Claim Coded

Trained coders.

"Nursing Progress Note 7p-7a\n\nNEURO: Sedated on Propofol gtt. Doesn't follow commands even when spoken in Greek by family member. Spont. non-purposeful movts noted. Easily aroused with touch, grimace and withdrawal of extremities noted. Sluggish pupils.\n\nCV: BP stable. HR 90s with occ. PACs. +murmur. Afebrile. +2 distal pulses. CVP 8-11. Warm extremities.\n\nRESP: Rate [**Month (only) **]. to 16 due to resp. alkalosis. Currently on CMV 40%/550/16/10. Scant secretions. Lungs essentially clear. No spont. breaths when resting.\n\nGI/GU: Abd. slightly firm and distended. No hemetemesis or melena. Currently NPO. Adequate UOP. No hematuria.\n\nSKIN: Yeast infection under skin folds. Excoriation of perianal area noted. Miconazole powder and nystatin cream used. Duoderm intact in sacral area.\n\nPLAN: Monitor Hct and transfuse if <30. SBT for possible extubation. Cont. octreotide gtt and manage BG with insulin gtt.\n"



Access Medicare (Cuatro LLC)

HEALTH INSURANCE CLAIM FORM

APPROVED BY NATIONAL UNIFORM CLAIM COMMITTEE (NUCC) 02/12

1. MEDICARE	
1a. INSURED'S I.D. NUMBER	ID1-12345
2. PATIENT'S NAME (Last Name, First Name, Middle Initial)	Doe, Jane
3. PATIENT'S BIRTH DATE SEX	01 16 09 M
4. INSURED'S NAME (Last Name, First Name, Middle Initial)	Doe, Jane
5. PATIENT'S ADDRESS (No., Street)	1000 Main Street
6. PATIENT RELATIONSHIP TO INSURED	Self
7. INSURED'S ADDRESS (No., Street)	1000 Main Street
CITY	Dallas
STATE	TX
ZIP CODE	75202
TELEPHONE (Include Area Code)	(832) 555-9876
9. OTHER INSURED'S NAME (Last Name, First Name, Middle Initial)	CareCore National LLC (Aetna Radiology)
10. IS PATIENT'S CONDITION RELATED TO:	
11. INSURED'S POLICY GROUP OR FECA NUMBER	GP1-67890
9a. OTHER INSURED'S POLICY OR GROUP NUMBER	ID2-12345
a. EMPLOYMENT? (Current or Previous)	NO
11a. INSURED'S DATE OF BIRTH SEX	01 16 09 M
b. AUTO ACCIDENT? PLACE (State)	NO
9b. RESERVED FOR NUCC USE	
c. OTHER ACCIDENT?	NO
9c. RESERVED FOR NUCC USE	
10d. CLAIM CODES (Designated by NUCC)	
9d. INSURANCE PLAN NAME OR PROGRAM NAME	
6. INSURANCE PLAN NAME OR PROGRAM NAME	GP2-67890
d. IS THERE ANOTHER HEALTH BENEFIT PLAN?	NO
12. PATIENT'S OR AUTHORIZED PERSON'S SIGNATURE	SIGNATURE ON FILE
DATE	2015-06-20
13. INSURED'S OR AUTHORIZED PERSON'S SIGNATURE	SIGNATURE ON FILE
14. DATE OF CURRENT ILLNESS, INJURY, or PREGNANCY (LMP)	
15. OTHER DATE	
16. DATES PATIENT UNABLE TO WORK IN CURRENT OCCUPATION	
17. NAME OF REFERRING PROVIDER OR OTHER SOURCE	
17a.	
17b. NPI	
18. HOSPITALIZATION DATES RELATED TO CURRENT SERVICES	
18. ADDITIONAL CLAIM INFORMATION (Designated by NUCC)	Memorial Hermann Imaging, 2900 Richmond Ave, Houston, TX, 77098
20. OUTSIDE LAB?	NO
21. DIAGNOSIS OR NATURE OF ILLNESS OR INJURY (Relate A-L to service line below (24E))	A. S02.91XA B. S06.899A
22. RESUBMISSION CODE	ORIGINAL REF. NO.
23. PRIOR AUTHORIZATION NUMBER	

24. A. DATE(S) OF SERVICE From MM DD YY	To MM DD YY	B. PLACE OF SERVICE	C. EMG	D. PROCEDURES, SERVICES, OR SUPPLIES (Explain Unusual Circumstances) CPT/HCPCS	MODIFIER	E. DIAGNOSIS POINTER	F. $ CHARGES	G. DAYS OR UNITS	H. EPSDT Family Plan	I. ID. QUAL.	J. RENDERING PROVIDER ID. #
1	06 20 15	06 20 15	41		ABK39	NH	A B	1500 00	1	NPI	1992733482
2	06 20 15	06 20 15	41		ABK35	NH	A B	178 40	22.3	NPI	1992733482
3	06 20 15	06 20 15	41		ABK22	NH	A B	2 20	1	NPI	1992733482
4	06 20 15	06 20 15	41		ABK24	NH	A B	750 00	3	NPI	1992733482
5										NPI	
6										NPI	

25. FEDERAL TAX I.D. NUMBER SSN EIN	65-4032840
26. PATIENT'S ACCOUNT NO.	AD0003F2
27. ACCEPT ASSIGNMENT? (For govt. claims, see back)	YES
28. TOTAL CHARGE	2430 60
29. AMOUNT PAID	5 00
30. Rsvd for NUCC Use	
31. SIGNATURE OF PHYSICIAN OR SUPPLIER INCLUDING DEGREES OR CREDENTIALS	
32. SERVICE FACILITY LOCATION INFORMATION	Courtyards At Pasadena 4948 Red Bluff Rd Pasadena, TX, 77503
33. BILLING PROVIDER INFO & PH #	(713) 456-7890
SIGNED	
DATE	
a.	
b.	
a.	1234567893
b.	

NUCC Instruction Manual available at: www.nucc.org

PLEASE PRINT OR TYPE

APPROVED OMB-0938-1197 FORM 1500 (02-12)

Coding Errors are Expensive

Errors in coding cost hospitals millions

- Claims Denials
- Incorrect base rate used for reimbursement

Coding Errors are Expensive

Denied claims

- Average $25 to reprocess
- 50-65% of denied claims are never worked

https://www.mgma.com/resources/revenue-cycle/you-might-be-losing-thousands-of-dollars-per-month

The Data

The Data

Data set from MIT

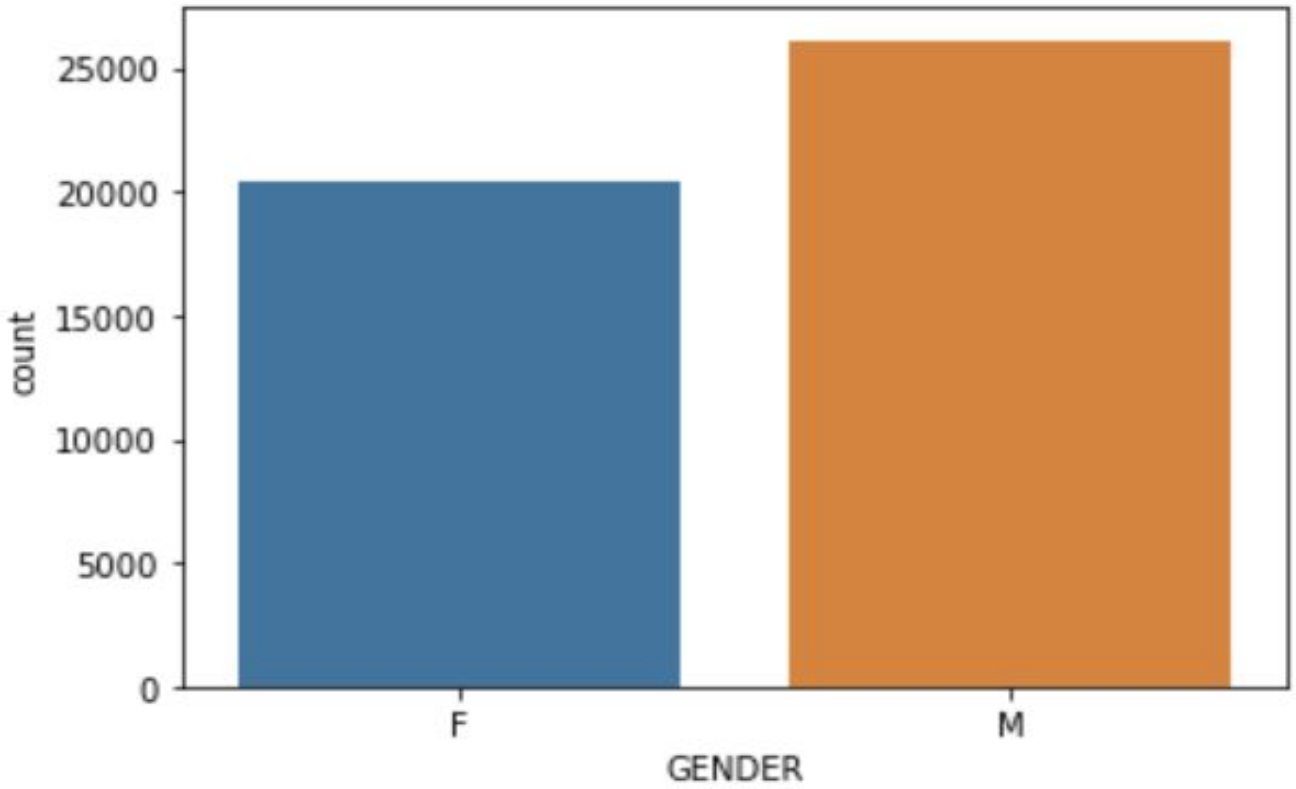
MIMIC-III

ICU patients at Beth Israel Deaconess Medical Center 2001-2012

Patient Data

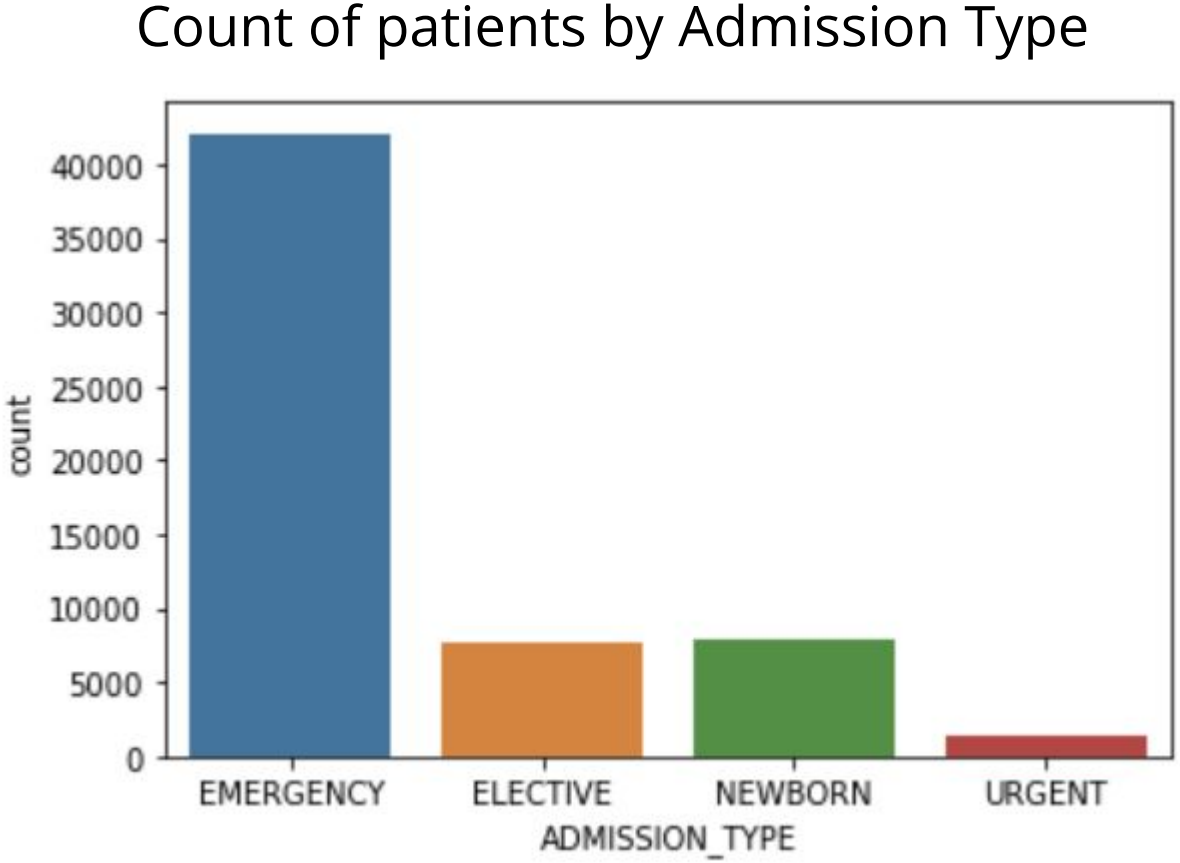
Over 40K
individual patients

Count of Male/Female patients



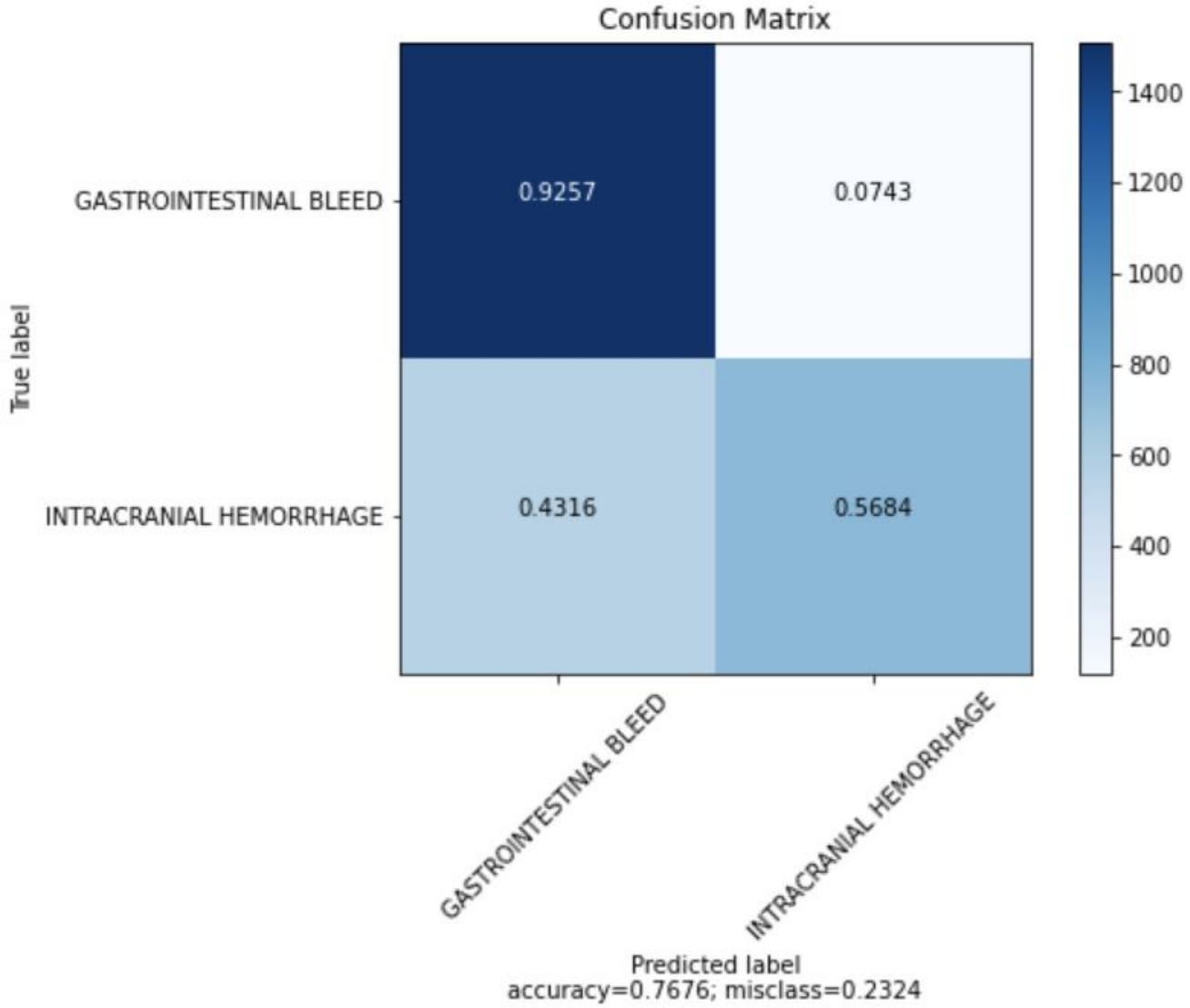
Patient Data

Majority admitted through the ED



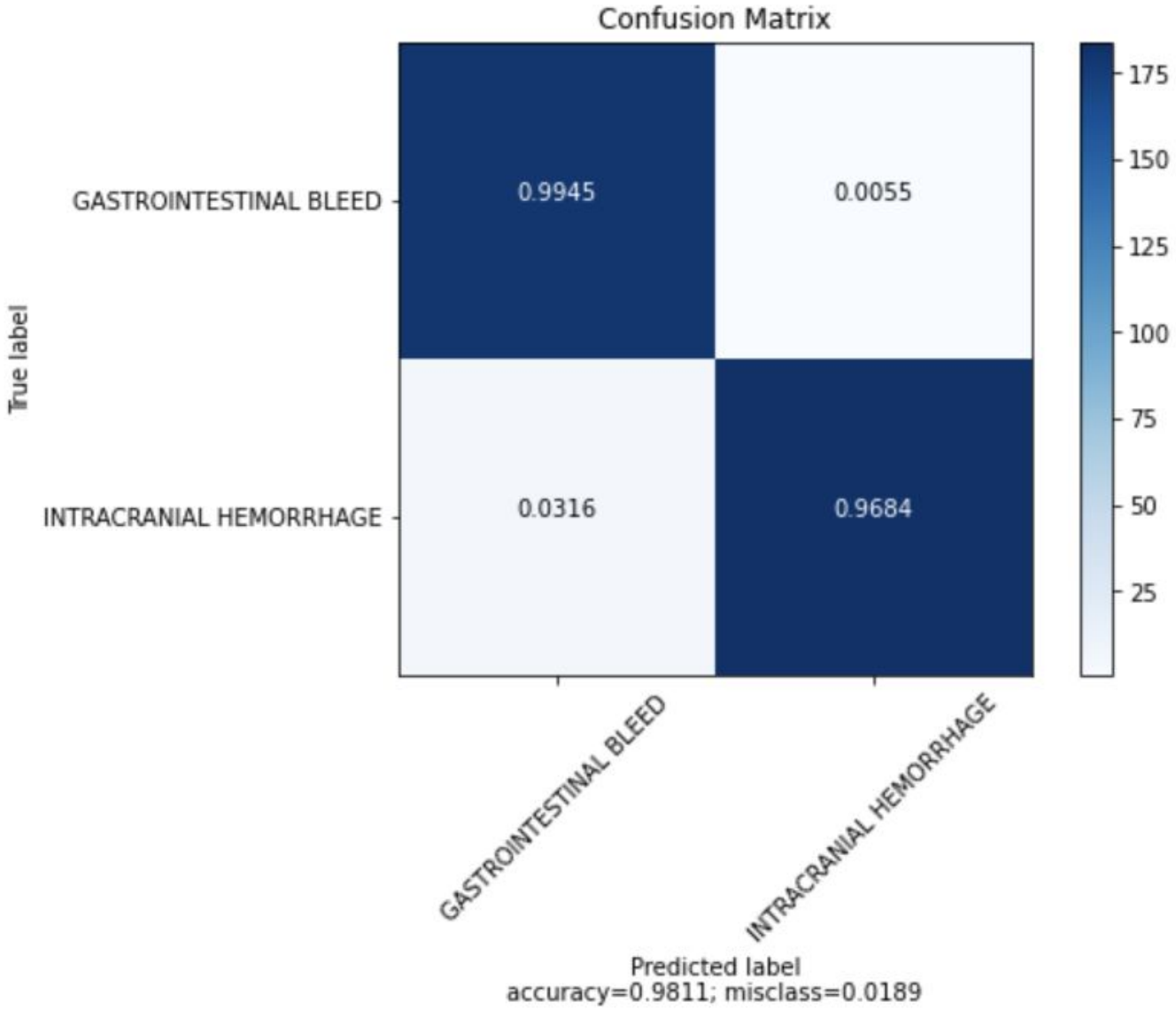
The Model

First Simple Model



- Chose 2 diagnoses
- Count vectorization
- Lemmatized
- 500 features
- Accuracy = .77

First Simple Model

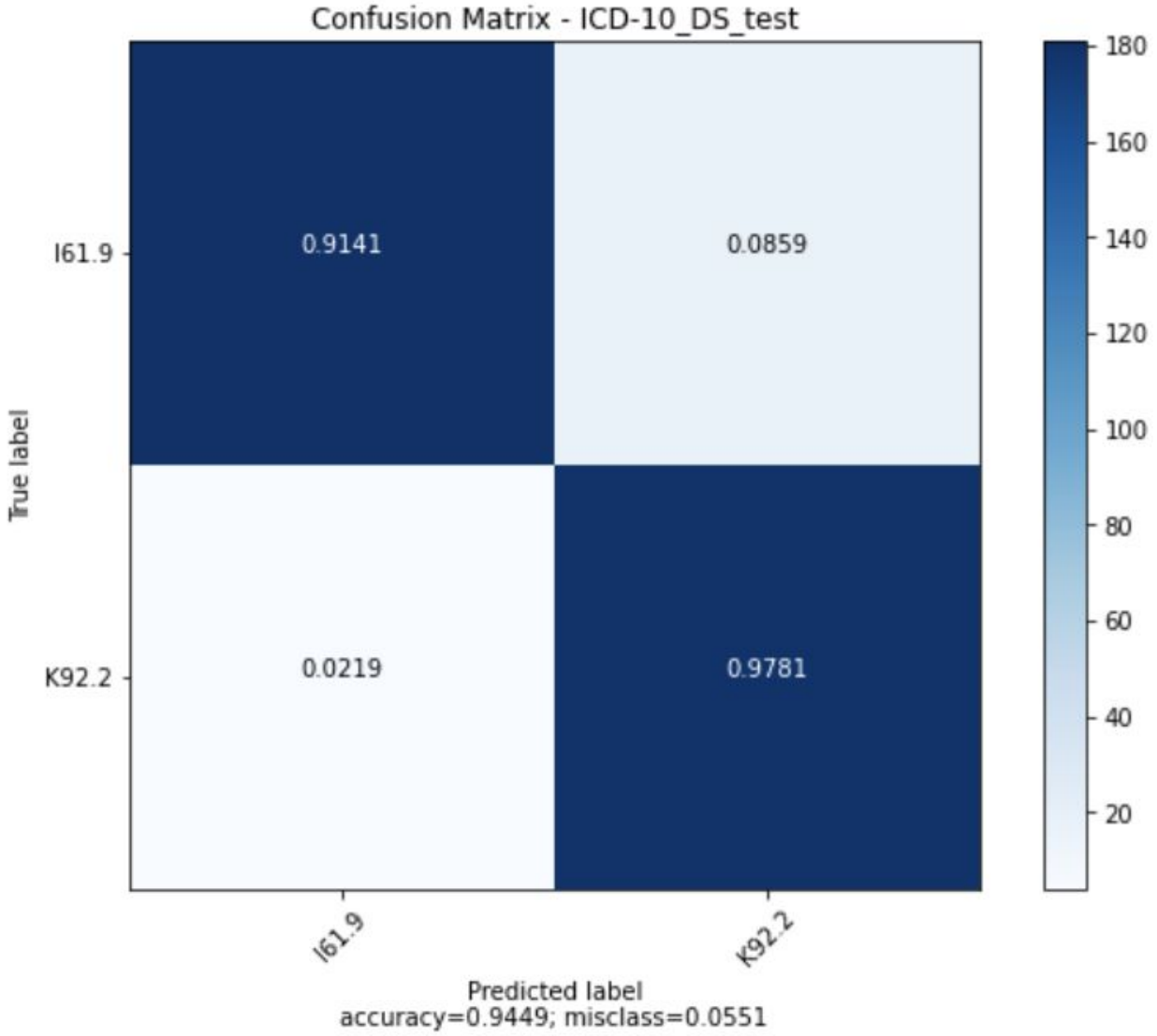


- Limited to Discharge Summary notes
- Accuracy = .98

Processed Text Included Actual Diagnosis

'admission date discharge date date birth sex f service medicine allergy topiramate aripiprazole shellfish bee pollen atten
laint am major surgical invasive procedure paracentesis thoracentesis history present illness f history alcoholic hepatitis
ge gi bleed without clear source subsequently develop c diff treat iv flagyl po vanc taper sbp hcap treat vanc cefepime dis
day transfer back hospital am report ems hospital rehab febrile today pcxr c w pna go svt rate transfer hospital give adeno
ive apap vanc zosyn transfer ed afebrile alter unable answer question labs show wbc stable hct plt electrolytes wnl excepti
ct p obtain look toxic megacolon negative however show large new right pleural effusion lead atelectasis near collapse righ
ill define opacity lul may reflect infectious process also splenomegaly varix signialying portal hypertension diagnositc pa
e inr low plt past ir requirement give iv flagyl possibility toxic megacolon also receive another l n ed sbps remain stable
t feel like truck hit elaborate much review system difficult obtain feel like truck hit sob abd pain distension past medica
sis p tip p cholecystectomy gastroesophageal reflux disease bipolar disorder htn depression anxiety recent burn hand housef
ocial history life husband child age smoke pack every week use accountant describes beer daily denies drug use family histo
ion exam hr bp l general encephalopathic mumble ox heent sclera icteric dry mm oropharynx clear eomi perrl neck supple jvp
l murmur rub gallops lung clear auscultation bilaterally anteriorly abdomen distend ascites present fluid wave diffusely te
l perfused pulse peripheral edema neuro cnii xii intact move extremity discharge exam v ra nr bmx x general chronically ill
e heent sclera icteric mmm cardiac tachycardic regular systolic murmur along left sternal border hyperdynamic precordium lu

Final Model



- Chose 2 ICD-10 codes
- Discharge Summary
- Count vectorization
- Lemmatized
- 500 features
- Accuracy = .95

Evaluation and Next Steps

Recommendations/Next Steps

- Proof of concept, NLP can be used to capture codes accurately
- Add additional diagnosis codes - medical coding is not binary
- Add in DRG - much lost revenue from incorrect DRG codes, specifically not capturing CC/MCC

Amanda Potter

amandaspotter@gmail.com

github.com/aspotter99

Contact