Automated Versus Manual ICD-10 Coding of Neuroimaging Reports ("Radnosis" Project)

Primary Category:
Informatics Innovations

Secondary Category:
Quality and Safety

Area of Focus:
Diagnostic Radiology

Author:

Jessica Moore, BS; MPH Student[1]

Co-Authors:

Alexander M. McKinney, MD[2,3]

Jeffrey Rykken, MD[2]

Bharathidasan Jagadeesan, MD[2]

Kevin D. Campbell, MHI[3,4]

Zeke J. McKinney, MD, MHI, MPH[1,3,5,6]

Affiliations:

1. University of Minnesota School of Public Health, Minneapolis, MN
2. University of Minnesota Department of Radiology, Minneapolis, MN
3. Veev, Inc., New Brighton, MN
4. Campbell Analytics, LLC, Minneapolis, MN
5. HealthPartners Occupational and Environmental Medicine Residency, St. Paul, MN
6. HealthPartners Institute, Bloomington, MN

Disclosures:

AM, ZM, KC are primary owners of Veev, Inc.

KC is the principal owner of Campbell Analytics, LLC.

No other relevant financial disclosures for the author or co-authors.

Background

- **2015: Implementation of ICD-10 in United States**
  - Significant impact on US billing and coding
- **Imaging Coding Requirements**
  - Include diagnostic codes affecting reimbursement
  - Typically radiologists do not aid in coding!
- **US Shift to Value-Based Care**
  - Billing and reimbursement to be linked to quality measures
  - Limited or no discrete data generated from radiology reports
  - How can **value** in radiology **be measured**?

Objective

This study aimed to test an automated algorithm's accuracy compared to manual coding for routine neuroimaging reports.

Methods: Demographics

- **Study Population**
  - Study date range: Jan. – Jun., 2016
  - 200 head, neck, and spine CT or MRI examination reports
  - Randomly-selected from 10,488 studies
  - 10,488 studies randomly-selected from larger set of studies performed during study time period
  - Equally-distributed from inpatient, outpatient, and ED

- **Study Site**
  - Academic health center serving most of Minnesota
  - Performs approx. 1 million imaging examinations per year

Methods: Coding

- **Automated Coding**
  - Experimental, proprietary natural language processing (NLP) engine
  - Extracted 645 "phrases" from 200 examination reports
    - 3.28 phrases per report
  - Coded phrases with multiple ICD-10 codes
  - Codes ranked by a score indicating likelihood of correctness

Methods: Coding

- **Manual Coding**
  - Simple web-based interface for phrase review
  - Select single "best" ICD-10 code per extracted phrase
  - Alternatively check "no applicable ICD-10 code exists" box

  - Four (4) physicians:
    - 3 neuroradiologists
    - 2 with content expertise in ICD-10 coding
  - Physicians blinded to NLP engine results

Web-Based Manual Coding Interface

Updated Phrase 4935 with Best Diagnosis 'Mass of left temporal lobe' (R22.0).

Impression Line	Phrase	Phrase ID	Best Dx Text	Best Dx ICD-10	Update/Delete Best Dx	ICD-10 N/A
2.3 CM HYPERDENSE MASS OF THE LEFT TEMPORAL LOBE WITH EXTENSIVE SURROUNDING VASOGENIC EDEMA. MILD ADJACENT LOSS OF THE GRAY/WHITE DISTINCTION IS QUESTIONED, WHICH COULD REFLECT SOME LOCAL PRESSURE-INDUCED ISCHEMIC CHANGE. DIFFERENTIAL CONSIDERATIONS OVERWHELMINGLY FAVOR TUMOR, EITHER METASTATIC OR PRIMARY BRAIN. MASS EFFECT FROM THE TUMOR AND ADJACENT VASOGENIC EDEMA IS RELATIVELY PRONOUNCED, WITH 8 MM OF LEFT-TO-RIGHT MIDLINE SHIFT AT THE LEVEL OF FORAMEN OF MONRO AND SLIGHT UNCAL HERNIATION.	hyperdense neoplasm left temporal lobe surrounding vasogenic edema	4935	Mass of left temporal lobe	R22.0	Delete 4935	☐
	mild adjacent loss gray white distinction questioned which could reflect some local pressure induced ischemia change	4936	Ischemic changes on head CT	I67.82	Delete 4936	☐
	differential considerations overwhelmingly favor neoplasm either metastasis primary brain	4937	Primary brain neoplasm (H)	D49.6	Delete 4937	☐
	compression neoplasm adjacent vasogenic edema relative pronounced 8 left right herniation level foramen monro slight uncal herniation	4938	Uncal herniation (H)	G93.5	Delete 4938	☐

Search by diagnoses OR by ICD-10 codes.

Search Diagnoses by Word (separate multiple words with space): mass temporal

Search Diagnoses by ICD-10 Code:

Found: **7 diagnoses matching text 'mass temporal'**

	Diagnosis Name	ICD-10	Diagnosis ID
○	Mass of petrous temporal bone	M89.9	198279
○	Right temporal lobe mass	R22.0	200208
○	Mass of right temporal lobe	R22.0	200209
○	Left parietotemporal mass	G93.9	215312
○	Mass of left parietotemporal region	G93.9	632488
◉	Mass of left temporal lobe	R22.0	635807
○	Mass of right parietotemporal region	G93.9	635810

Methods: Manual Coding Analysis

- **Agreement of Applicable ICD-10 Codes**
  - Fleiss' κ (kappa): agreement between ≥ 3 raters
- **Agreement of "Best" ICD-10 Code Among Reviewers**
  - % agreement between combinations of 2-4 reviewers
  - Phrases where combination of reviewers assigned ICD-10 code
  - Varied for length of ICD-10 code (3, 4, 7 characters)

Methods: Coding Length

- **ICD-10 Code Length**
  - Maximum 7 characters (excluding period)
  - "Cerebral infarction due to thrombosis of left MCA" (I63.312)

ICD-10 Character Length	Character Specification	Example
3	General pathology	"I63"
4	Specific pathology	"I63.3"
5	Anatomic localization	"I63.31"
6	Detailed specification, e.g., laterality	"I63.612"
7	Initial or subsequent encounter, sequela	"I63.612A"

Methods: Automated Coding Analysis

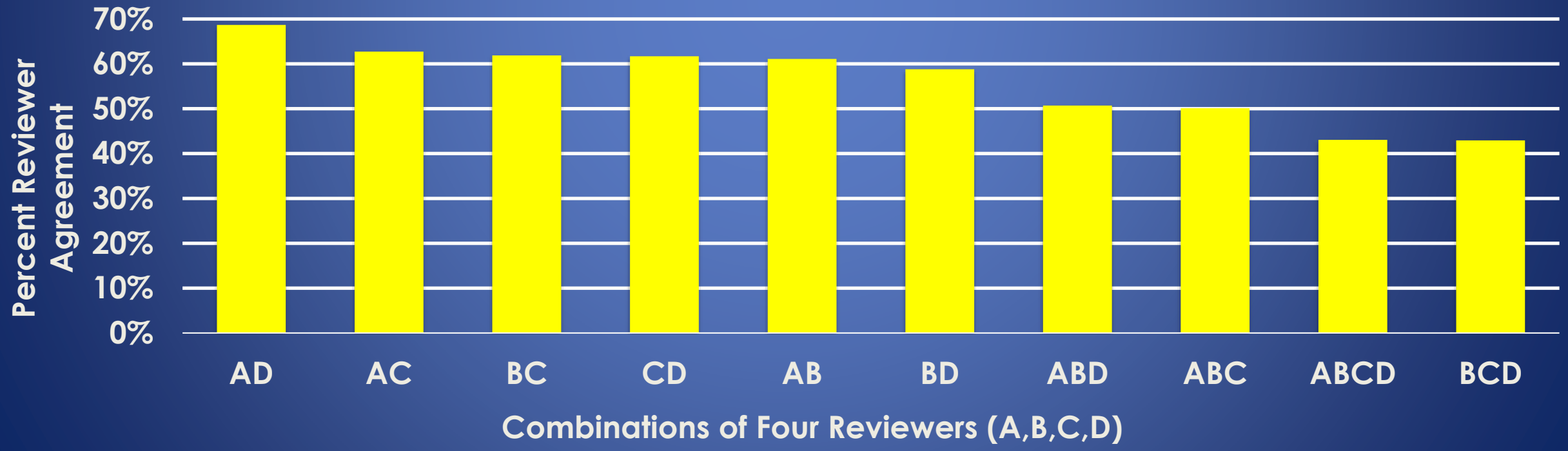
- **Agreement of "Best" ICD-10 Code with NLP Engine**
  - Evaluation for engine matches with all 4 reviewers
  - Varied for length of ICD-10 code (3, 4, 7 characters)
  - Varied for number of ranked engine ICD-10 matches
    - e.g., top 1, top 5, top 10
  - Threshold engine score:
    - Average likelihood score for the top ranked matches where reviewers and engine agreed at a specified ICD-10 code length

Classification	Reviewers	Engine
True Positive (TP)	All agreed	Agreed with reviewers
False Negative (FN)	All agreed	Did not agree with reviewers
True Negative (TN)	All not agreed	Score < threshold
False Positive	All not agreed	Score ≥ threshold

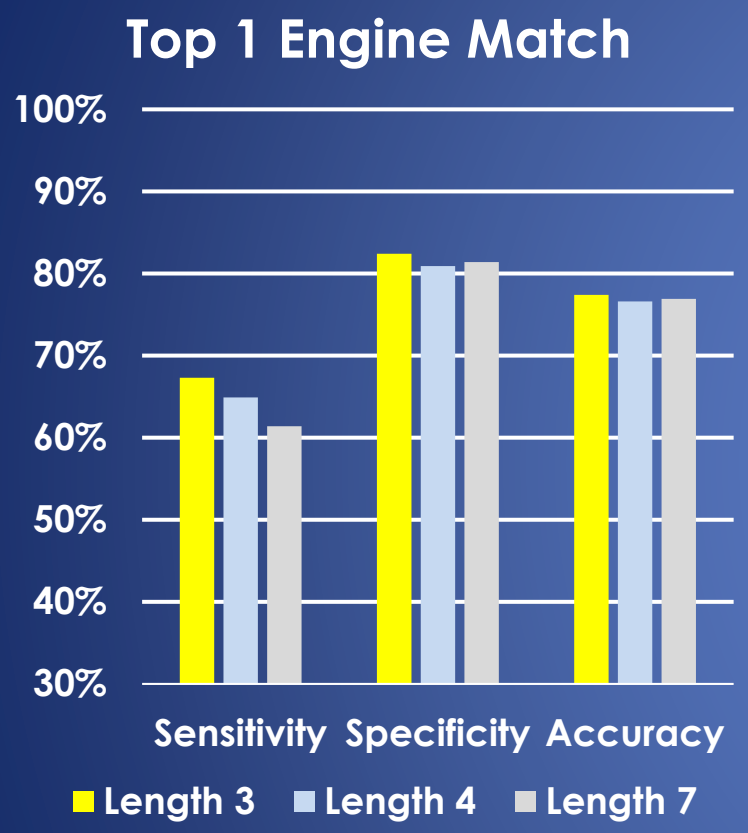
Results: Reviewer Agreement

- **Agreement of an Applicable ICD-10 Code**
  - Fleiss' κ (kappa) = 0.546 (moderate agreement, $p << 0.05$)

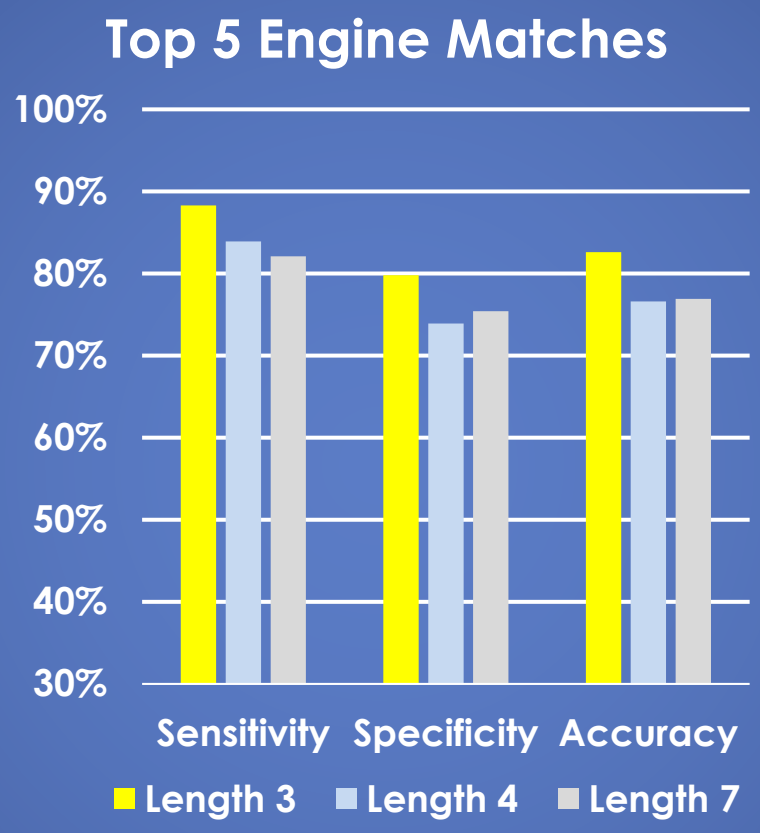
- **Agreement of "Best" ICD-10 Code Among Reviewers**



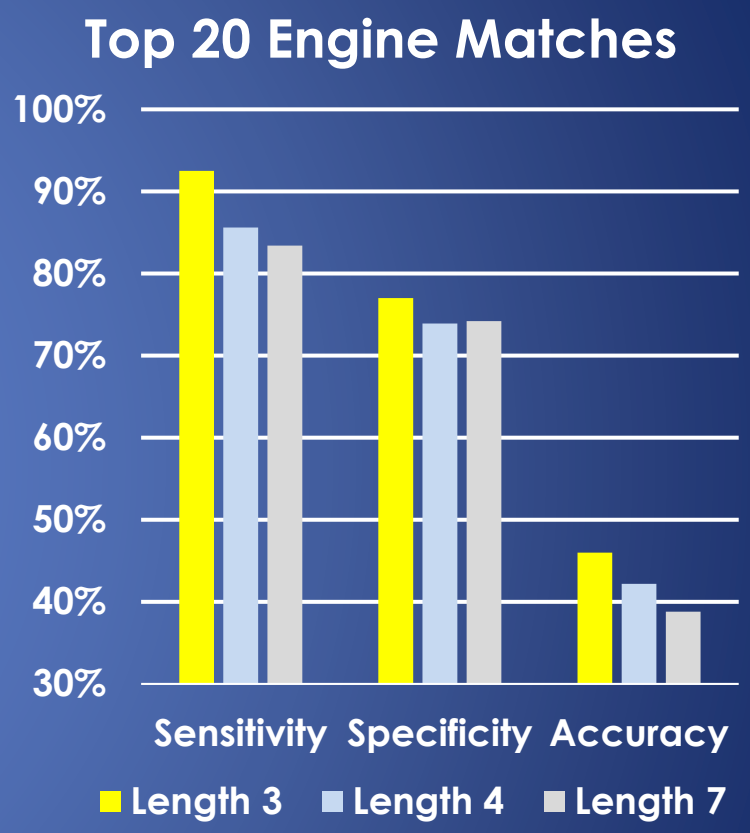
Results: Engine Performance



OR: 9.6 7.8 7.0



OR: 29.9 14.8 14.0



OR: 41.5 16.9 14.5

Performance metrics:

Sensitivity (recall): TP / (TP + FN)

Accuracy: (TP + FP) / n

n = number of phrases (645)

Specificity: TN / (TN + FP)

Odds Ratio (OR): (TP / FP) / (FN / TN)

Conclusions

- **Agreement in Manual Coding**
  - Limited agreement on whether ICD-10 codes can be applied
  - Limited agreement on assignment of ICD-10 codes

- **Agreement with Automated Coding**
  - NLP-based coding algorithm **highly sensitive for top 5 codes**
  - Increasing to top 20 codes **decreases** performance

  Note: Having multiple codes attached to a report is relevant
  - as up to 4 can be attached to billing

Conclusions

- **Implications**
  - Coding variability among expert reviewers
  - Need to improve coding homogeneity
  - Automated coding-based decision support for radiologists during dictation ("real-time") can aid in improving coding homogeneity