Automated Classification of Cancer Pathology Reports

George Cernile
Artificial Intelligence in Medicine, Inc
Eric B. Durbin, MS
Kentucky Cancer Registry
NAACCR 2011 Conference
Louisville Kentucky, USA
June 23, 2011

Introduction

- Problem
- Approach
- Results
- Conclusion

Background

- Volume of electronic pathology (E-Path) reports at central cancer registries increasing rapidly
- Key elements in E-Path reports reside in narrative text
- E-Path utility is limited by staff resources available to manually review large volumes of reports
- Reliable automated coding (auto-coding) of even minimal diagnostic elements would improve efficiency and utility

Reliable Coding of key data elements in E-Path reports would enhance:

- Rapid case ascertainment
- Case level linkages
- Case finding
- Increased abstracting efficiencies
- Audits
- Early assessments of incidence rates

The Problem

- Currently done manually - requires review of path reports – takes time
- Automating would save labor, improve data acquisition
- Requires trained registrar to perform task
- Trained registrars may not always agree on the final answer
- Automated system needs to perform to 95% accuracy to be effective

The problem

- E-Path produces ICD-O-3 codes
  - ICD-O-3 codes may not be enough (too many, not the right ones)
  - Other tumor information may be used to interpret site and procedure
- Synoptex produces a checklist of 15-20 question/answer pairs using NLP

- Goal: Target fewer critical/key data elements such as
  - SEER Site Groups (Site+Histology)
  - Behavior
  - Laterality
  - Grade

Example1 – ICD-0-3 codes

M-80103

M-85003

M-80003

C50.9

C77.9

CLINICAL HISTORY/MACROSCOPY

Right mastectomy and axillary tissue. A right mastectomy specimen with overlying skin measuring 220mm x 85mm and underlying breast tissue measuring 220mm x 100mm x 70mm. The axillary tail measures 125 x 60mm. The nipple is slightly retracted and located centrally. The superior margin is painted red, the inferior margin painted green and the deep cut margin is painted blue. Cut sections of the underlying breast tissue shows a ill-defined grey white yellow lesion with patchy areas of haemorrhage measuring 35 x 35 x 35mm located immediately below the nipple, 20mm from the inferior margin, 45mm from the deep cut margin, 50mm from the superior margin, 85mm from the medial margin and 100mm from the lateral cut margin. A1 - nipple, B1 - upper outer quadrant, C1 - upper inner quadrant, D1 - lower outer quadrant, E1 - lower inner quadrant, F1, G1 - tumour composite blocks, H1, I1 - tumour composite blocks, J1 - deep cut margin, K1 - superior margin, L1 - inferior margin, M4 - lymph nodes, N4 - lymph nodes, O - 3 serial slices lymph node, P - 3 lymph nodes.

MICROSCOPY

This right mastectomy specimen demonstrates an invasive ductal carcinoma with the following pathological features:

TUMOUR HISTOLOGY & GRADE

The tumour is of an infiltrating poorly differentiated ductal carcinoma of non-otherwise specified type. The tumour is poorly defined and extremely infiltrative, comprising poorly-formed tubules, nests or strands of cuboidal tumour cells displaying high grade nuclei. The tumour cells are set within fibrotic desmoplastic stroma. Many lactiferous ducts are entrapped within the tumour. Frequent tumour mitoses are seen. Microcalcification is seen in some neoplastic tubules.

Tumour grade (Modified Bloom-Richardson Scoring System):

Tubular formation:	3
Nuclear atypia:	3
Tumour mitoses:	2
Total score:	8 (Grade III)

Coding Accuracy Reference Builder - Multiple Files

File Update Reference Tools

☑	600140...	IGNOR
☑	FWDN...	MTLG
☑	FWDNO...	MTLG
☑	CPAC...	MLG
☑	CPACP...	MTLG
☑	KY_AP-...	G
☑	KY_A...	MTLG
☑	KY_A...	TLG
☑	KY_A...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_A...	TLG
☑	KY_AP-...	MTLG
☑	KY_A...	LG
☑	KY_A...	MLG
☑	KY_A...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP-...	LG
☑	KY_AP-...	MTLG
☑	KY_AP-...	TLG
☑	KY_AP-...	L
☑	KY_AP-...	MTLG
☑	KY_AP-...	MTLG
☑	KY_AP...	MTLG

[OTHER]

NONE AVAILABLE.

TEXT DIAGNOSIS FROM PRIOR CASES:
KC08-48 Accsnd 09JAN2008, Signed Out 24JAN2008, Pathologist CTA
AMENDED DIAGNOSIS - SPECIMEN SOURCE

RIGHT BRONCHIAL **WASH**:
-NON-SMALL CELL CARCINOMA.
-INTRADEPARTMENTAL CONSULTATION WITH AGREEMENT IS OBTAINED.

CTA/erm
KC08-30 Accsnd 08JAN2008, Signed Out 09JAN2008, Pathologist SPIRE
RIGHT LUNG, BRONCHIAL BRUSHING, THIN LAYER PREPARATION AND CELL BLOCK:
-POSITIVE FOR MALIGNANCY,
KERATINIZING SQUAMOUS CELL CARCINOMA PRESENT.

COMMENT: Correlation with biopsy is confirmatory of the diagnosis.
A) Bronchial Brushing
Tissue Category source: Respiratory
Right bronchial brushing. Two thin preps. Cell block prepared. Specimen received fixed in Cytolyt. 35 ml. Mucoid specimen. Tissue fragments present. Routine AFB-fungus. C-4-08
jck
The tumor is notable for necrosis, numerous malignant squamous cells present, notable for glassy eosinophilic cytoplasm. Intradepartmental consultation with agreement is obtained.
SES/erm
RIGHT LUNG, BRONCHIAL BRUSHING, THIN LAYER PREPARATION AND CELL BLOCK:
-POSITIVE FOR MALIGNANCY,
KERATINIZING SQUAMOUS CELL CARCINOMA PRESENT.

Automatic Coding

Criteria	Value
Morphology	M-80463
Topography	C34.9
Laterality	1
Grade	9

Manual Coding

Criteria	Value
Morphology	M-80713
Topography	C34.9
Laterality	1 Right origin
Grade	9 Unknown or N/A

☑ Reviewed

Comments

Morphology: upon review coder changed to Keratinizing squamous cell carcinoma 80713. There are actually two specimens noted within this report, one of which has a diagnosis of non-small cell carcinoma (80463), however, the more specific diagnosis of keratinizing squamous cell carcinoma would be used to code the case.

Raw Output

Criteria	Value
Morphology	M-80103
Morphology	M-80413
Morphology	M-80463
Morphology	M-80703
Morphology	M-80713
Morphology	M-80001
Morphology	M-80011
Morphology	M-80013
Topography	C34.9
Laterality	1
Grade	9

C# | CLips

CODEX Explanation

Priority 4: C Code in Associated Link

M-80463 C34.9: Priority 2

Associated Link Information:
****

Step 8: Remove Same Family CCode in order C**.9 which means if Cxy.z and Cxy.9 then remove Cxy.9, Occurence and no Associate Link:

Associated Links Informations:

****

Step 9: Remove M code behavior 0/1 if have M code behavior 2/3:

****

Find :

Processing Reports...

Knowledge base examples

Reduction rules (general/specific)

- If the report contains a C42(0/1/2) (Blood/Bone Marrow/Spleen) then assume this is a Leukemia report and remove codes that are non Leukemia morphologies

- IF M-83233 Mixed cell adenocarcinoma THEN REMOVE M-81403 Adenocarcinoma NOS

Deduction rules to add codes

- DEDUCE M-85223 Infiltrating duct and lobular carcinoma (C50._) IF M-82303 Solid adenocarcinoma with mucin formation AND M-85003 Duct adenocarcinoma NOS

Deduction rules – adding codes

- DEDUCE M-85223 Infiltrating duct and lobular carcinoma (C50._) IF M-85203 Infiltrating lobular carcinoma (C50._) AND M-85003 Duct adenocarcinoma NOS

- DEDUCE M-85223 Infiltrating duct and lobular carcinoma (C50._) IF M-82303 Solid adenocarcinoma with mucin formation AND M-85003 Duct adenocarcinoma NOS

Other rules

- Statistical reference for code pairing
- Direct rules from reviewer feedback

Remove C code in the same segment
if there is no M code and C code pair in the same segment
if and only if the C Code has no link with the M Code
according to the LinkCode table

Knowledge Acquisition

'FWDNORTONC0000008467'	Morphology: Diagnosis states endocrine carcinoma. This terminology does not exist in ICD-O-3. Consensus opinion is 8246 neuroendocrine carcinoma. this is not certain.
'FWDNORTONC0000008475'	Morphology: adenocarcinoma. code 81403. Do not use 81443 unless stated as "intestinal type".
'FWDNORTONC0000008521'	Morphology: diagnosis states adenocarcinoma in 6a. 81403. do not code 8144 unless stated as adenocarcinoma, intestinal type.
'FWDNORTONC0000008526'	Morphology: Diagnosis 5c states DCIS is solid, cribriform and comedo TYPES. 85232 Using 2007 multiple primary and histology rule H6, if there is intraductal carcinoma and tow or more specific intraductal types from table 3 pg 50, then use 85232.

Knowledge Acquisition

'FWDNORTONC0000008467'	Morphology: Diagnosis states endocrine carcinoma. This terminology does not exist in ICD-O-3. Consensus opinion is 8246 neuroendocrine carcinoma. this is not certain.
'FWDNORTONC0000008475'	Morphology: adenocarcinoma. code 81403. Do not use 81443 unless stated as "intestinal type".
'FWDNORTONC0000008521'	Morphology: diagnosis states adenocarcinoma in 6a. 81403. do not code 8144 unless stated as adenocarcinoma, intestinal type.
'FWDNORTONC0000008526'	Morphology: Diagnosis 5c states DCIS is solid, cribriform and comedo TYPES. 85232 Using 2007 multiple primary and histology rule H6, if there is intraductal carcinoma and tow or more specific intraductal types from table 3 pg 50, then use 85232.

Knowledge Acquisition

'FWDOMHSOMH0000008119'	Morphology: when both invasive and insitu carcinoma is present in same breast specimen use behavior code 3
'FWDUKYUKY000000488 69'	Morphology: use diagnosis metastatic squamous cell carcinoma 80703. A prior case is referenced below with the 80723 code, but this specimen does not have that specific diagnosis.
'ULEPATH000000003450'	Topography: Comment section states clinical findings FAVOR lung primary with metastasis to the small bowel. Using the companion document which follows ULEPATH0000003451 the site of the primary tumor can be determined to be left lower lobe lung c34.3 laterality 1. This information does not appear in this report. Both reports would be read together as they are from the same autopsy procedure.

Knowledge Acquisition

'FWDOMHSOMH0000008119'	Morphology: when both invasive and insitu carcinoma is present in same breast specimen use behavior code 3
'FWDUKYUKY000000488 69'	Morphology: use diagnosis metastatic squamous cell carcinoma 80703. A prior case is referenced below with the 80723 code, but this specimen does not have that specific diagnosis.
'ULEPATH000 00000 03450'	Topography: Comment section states clinical findings FAVOR lung primary with metastasis to the small bowel. Using the companion document which follows ULEPATH0000003451 the site of the primary tumor can be determined to be left lower lobe lung c34.3 laterality 1. This information does not appear in this report. Both reports would be read together as they are from the same autopsy procedure.

New Approach

Lessons from Synoptex

- Uses all information in path report to determine site and procedure, not only ICD-O-3 codes
- More accurate determination of SEER site buckets

Some AI concepts

ICD-0-3 codes processing has limits. Need more information – use all information in path report, not only ICD-O-3 codes

Synoptex approach

- Generate all possible data for each checklist, assign weighted scoring, (eg: mastectomy – high from Breast)
- Test and reduce until one site is produces.

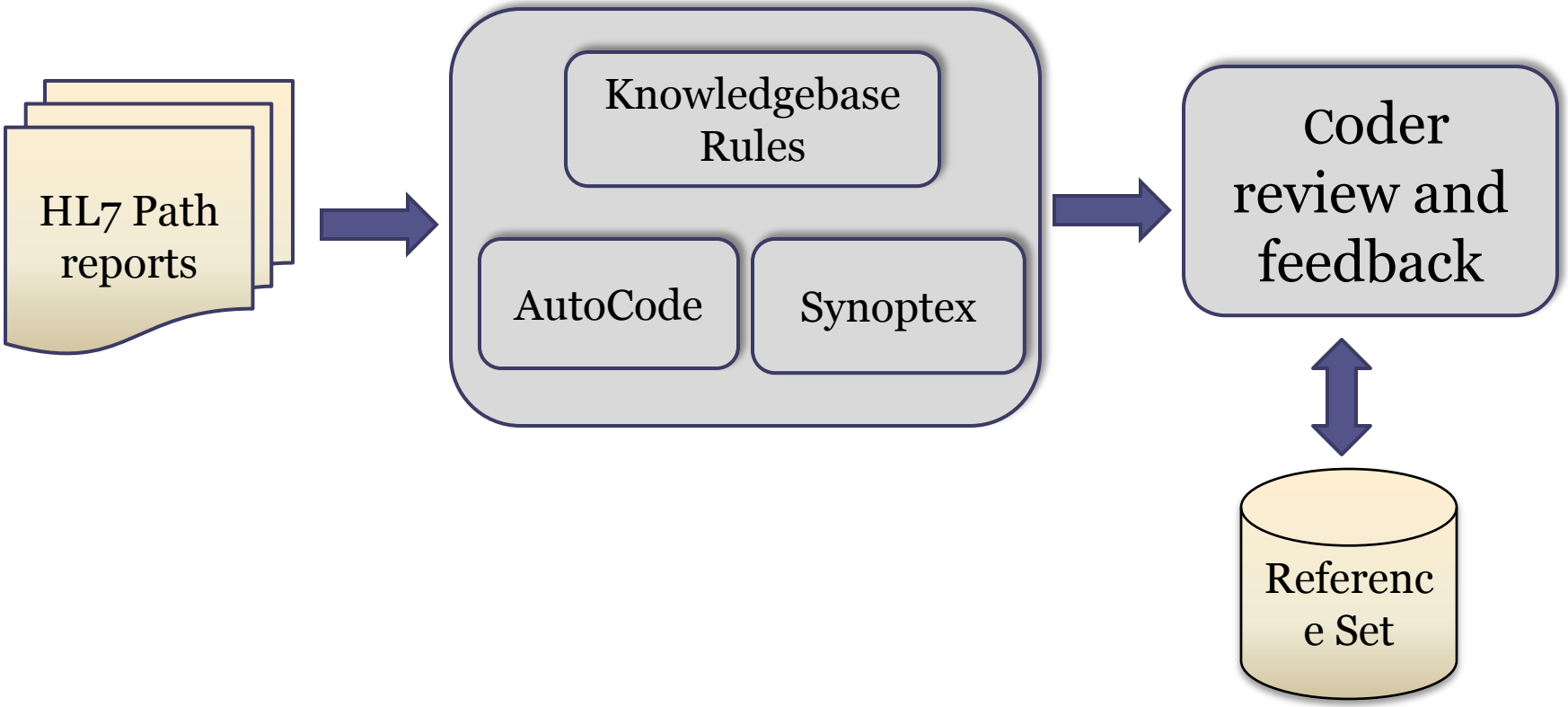
Uses forward chaining because we have all data upfront in path report. As opposed to backward chaining looks up data as needed –eg: Diagnosis…

Example – Generate and Test

Adenocarcinoma: score 1 for every site
Her2: score 10 for breast
Mastectomy: score 10 for breast
Laterality: score 1 for all paired sites
…
For each report can generate up to 200-400 items

Final score: Breast 22, Lung 2, Other sites 1
Remove all non Breast data
Apply heuristics to continue data reduction

Codex System



Measuring System performance

- Kentucky Cancer Registry provided expertise for knowledge acquisition
- Three professional coders to evaluate Codex performance – formed consensus panel
- Codex was measured against consensus panel results

Inter-Coder Reliability (All Coders Agree) (N=600)

Reportable	494	82.3%
SEER Site	469	78.2%
Histology	441	73.5%
Behavior	503	83.8%
Topography	384	64.0%
Topography (Major Site)	44	73.7%
Laterality	451	75.2%
Grade	439	73.2%

Codex Overall Accuracy Compared to Consensus Panel(N=393)

Reportable	366	93.1
SEER Site	280	71.2
Histology	252	64.1
Behavior	343	87.3
Topography	238	60.6
Topography (Major Site)	296	75.3
Laterality	285	72.5
Grade	286	72.8

Codex Accuracy for Breast Compared to Consensus Panel (N=64)

Reportable	64	100.0
SEER Site	58	90.6
Histology	42	65.6
Behavior	57	89.1
Topography	45	70.3
Topography (Major Site)	58	90.6
Laterality	55	85.9
Grade	59	92.2

Codex Accuracy for Colon and Rectum Compared to Consensus Panel (N=37)

Reportable	35	94.6
SEER Site	31	83.8
Histology	25	67.6
Behavior	32	86.5
Topography	14	37.8
Topography (Major Site)	25	67.6
Laterality	28	75.7
Grade	25	67.6

Codex Accuracy for Lung Compared to Consensus Panel (N=43)

Reportable	43	100.0
SEER Site	35	81.4
Histology	26	60.5
Behavior	43	100.0
Topography	28	65.1
Topography (Major Site)	35	81.4
Laterality	30	69.8
Grade	40	93.0

Codex Accuracy for Prostate Compared to Consensus Panel (N=15)

Reportable	15	100.0
SEER Site	14	93.3
Histology	12	80.0
Behavior	13	86.7
Topography	15	100.0
Topography (Major Site)	15	100.0
Laterality	15	100.0
Grade	12	80.0

Codex Accuracy for Kidney and Renal Pelvis Compared to Consensus Panel (N=10)

Reportable	10	100.0
SEER Site	10	100.0
Histology	5	50.0
Behavior	10	100.0
Topography	10	100.0
Topography (Major Site)	10	100.0
Laterality	10	100.0
Grade	7	70.0

Conclusion

- System shows improvements as more knowledge is added
- Can be used for the more common sites – improve accuracy for these sites
- May need new approaches to knowledge acquisition –
  - Direct user input
  - Automated decision trees/Rules

Acknowledgements

Funding

- ◦ NCI/SEER Program: N01-PC-54403

CTR Coders

- ◦ Lisa Witt
- ◦ Kelly Pictor
- ◦ Elaine Collins

KCR Data Management

- ◦ Tamas Gal
- ◦Isaac Hands

AIM Development Team