Leveraging Unstructured Data for EHR-based Prediction

Matt Engelhard

Lecture Outline

- What is unstructured data?
- Why would we want to use unstructured data?
- How do we use unstructured data (in isolation) to make predictions?
- How do we use unstructured data and structured data together to make predictions?

What is unstructured data?

Structured Data

Definition: data organized into specific, queryable fields as part of an overarching model, with each field having a defined purpose.

- typically stored in tabular (i.e. table, spreadsheet) format
- numeric variables, categorical variables and/or short free text

Examples: demographics, diagnoses, procedures, labs, meds, vitals

Example from Duke's Clinical Research Datamart…

PCOR_RSH_DIAGNOSIS_VW EXAMPLE

DIAGNOSISID	PATID	ENCOUNTERID	ENC_TYPE	ADMIT_DATE	PROVIDERID	DX	DX_TYPE	DX_SOURCE	DX_ORIGIN	PDX	DX_POA	RAW_DX	RAW_DX_TYPE	RAW_DX_SOURCE	RAW_PDX
36510180	123456	57896636	AV	2014/02/21 00:00:00	1745326	530.9	9	FI	BI	OT	NI	530.9	ICD-9-CM	FINAL DIAGNOSIS	PRIMARY
36510407	223456	57897312	AV	2014/05/07 00:00:00	73411	530.81	9	FI	BI	OT	NI	530.81	ICD-9-CM	FINAL DIAGNOSIS	PRIMARY
36510405	323456	57894371	ED	2014/06/13 00:00:00	123901	553.3	9	FI	BI	OT	NI	553.3	ICD-9-CM	FINAL DIAGNOSIS	SECONDARY
26510406	423456	57843289	AV	2015/01/21 00:00:00	1719263	530.3	9	FI	BI	OT	NI	530.3	ICD-9-CM	FINAL DIAGNOSIS	SECONDARY
36510721	523456	57898224	AV	2015/03/04 00:00:00	1719263	787.2	9	FI	BI	OT	NI	787.2	ICD-9-CM	FINAL DIAGNOSIS	PRIMARY
36510720	612345	57899120	ED	2015/07/30 00:00:00	156342	454.8	9	FI	BI	OT	NI	454.8	ICD-9-CM	FINAL DIAGNOSIS	PRIMARY
36511049	654321	57892268	EI	2016/12/03 00:00:00	165493	V12.72	9	FI	BI	OT	NI	V12.72	ICD-9-CM	FINAL DIAGNOSIS	SECONDARY
36508682	543216	57893648	IP	2016/08/05 00:00:00	178238	211.3	9	FI	BI	OT	NI	211.3	ICD-9-CM	FINAL DIAGNOSIS	PRIMARY

COLUMN	DEFINITION
DIAGNOSISID:	An arbitrary identifier for each record. Not currently useful for linkage across tables.
ENCOUNTERID:	The ENCOUNTERID (encounter ID) is an arbitrary identifier for a specific encounter that can be used to link information across tables in the model.
PATID:	The PATID (patient identifier) is an arbitrary, person-level identifier used to link information across tables within the data model. There is a table within the CRDM that allows an investigator with an appropriate IRB protocol to link PATID to MRN.
ENC_TYPE:	Encounter type indicates the setting in which the encounter took place (the same data are provided in the encounter table: AV=Ambulatory Visit, ED=Emergency Department, EI=Emergency Department Admit to Inpatient Hospital Stay, IP=Inpatient Hospital Stay, OS=Observation Stay, NI=No information, UN=Unknown, OT=Other
ADMIT_DATE:	The date on which the encounter began (also included in the encounter table).
PROVIDERID:	Unique identifier for a provider ? can be used to link an encounter to the provider table.
DX:	The diagnosis code; uses either ICD-9 or ICD-10
DX_TYPE:	09=ICD-9-CM, 10=ICD-10-CM, 11=ICD-11-CM, SM=SNOMED CT, NI=No information,

Unstructured Data

Definition: data that is NOT structured data

- in other words, everything else
- typically stored in a modality-specific format

Examples:
- clinical notes
- imaging (e.g. radiology, ophthalmology, ultrasound, pathology)
- physiologic time series (e.g. electrophysiologic studies, digital health)

Why use unstructured data?

A patient presents with shortness of breath…

- Presents to urgent care and is transferred to the ED

- Evaluated in the ED
  - History of present illness
  - Review of medical history (including meds and allergies) + social and family hx
  - Physical exam and review of systems
  - Lab results: BMP and CBC
  - Imaging & other studies: CT and EKG

- Admitted to the hospital

What will we find in the structured data?

- Encounter details (ED admission)
- Labs (BMP, CBC)
- Vital signs
- Procedure codes (CT)
- Diagnosis codes (PE)

- …what *won't* we find?

The H&P

written by a UNC student and publicly available:

https://www.med.unc.edu/medclerk/education/grading/history-and-physical-examination-h-p-examples/

FHIR systems **SHALL** support, at *minimum*, these eight "Common Clinical Notes":

1. Consultation Note (11488-4)
2. Discharge Summary (18842-5)
3. History & Physical Note (34117-2)
4. Procedures Note (28570-0)
5. Progress Note (11506-3)
6. Imaging Narrative (18748-4)
7. Laboratory Report Narrative (11502-2)
8. Pathology Report Narrative (11526-1)

Chief Complaint:

Shortness of breath.

History of the Present Illness:

Mr. [redacted] is a previously healthy 56-year-old gentleman who presents with a four day history of shortness of breath, hemoptysis, and right-sided chest pain. He works as a truck driver, and the symptoms began four days prior to admission, while he was in Jackson, MS. He drove from Jackson to Abilene, TX, the day after the symptoms began, where worsening of his dyspnea and pain prompted him to go to the emergency room. There, he was diagnosed with pneumonia and placed on Levaquin 500 mg daily and Benzonatate 200 mg TID, which he has been taking for two days with only slight improvement. He then drove from Abilene back to Greensboro, where he resides, and continued to experience shortness of breath, right sided chest pain, and hemoptysis. He presented to an urgent care office in town today, and was subsequently transferred to the Moses Cone ER due to the provider's suspicion of PE.

The right-sided pain is located midway down his ribcage, below the axilla. This pain is sharp, about 7/10 in severity, and worsens with movement and cough. Pressing on the chest does not recreate the pain. He feels that the pain has improved somewhat over the past two days. The hemoptysis has been unchanged since it began; there is not frank blood, but his sputum has been consistently blood-tinged. The blood seems redder at night. The dyspnea has been severe, and it is difficult for him to walk more than across a room. He states that he feels as though there is a "rattling" in his chest. At baseline, he experiences no dyspnea on exertion and has no history of COPD or other respiratory problem. He is a smoker, smoking a little less than a pack a day for thirty-five years. Past history is notable for the fact that he experienced transient left lower leg swelling – from below the knee down – and pain several weeks ago during a cross-country haul. He also notes a four day history of decreased appetite, poor sleep, and subjective fever and chills, with a measured fever of 103 in the hospital in Abilene. He had a bout of pneumonia about two months ago, but has been healthy for the most part and denies any chronic medical conditions. Currently he is fairly comfortable, with morphine helping with the pain. He has no history of a clotting disorder, no cardiac history, and denies any chest trauma or aspiration. He has had no sick contacts.

The H&P

written by a UNC student and publicly available:

https://www.med.unc.edu/medclerk/education/grading/history-and-physical-examination-h-p-examples/

FHIR systems **SHALL** support, at *minimum*, these eight "Common Clinical Notes":

1. Consultation Note (11488-4)
2. Discharge Summary (18842-5)
3. History & Physical Note (34117-2)
4. Procedures Note (28570-0)
5. Progress Note (11506-3)
6. Imaging Narrative (18748-4)
7. Laboratory Report Narrative (11502-2)
8. Pathology Report Narrative (11526-1)

Past Medical History:

1. Hernia repair
2. Bilateral thumb surgeries, secondary to two separate injuries sustained while working with machinery

Medications:

No regular medications, over-the-counter medications, or supplements. Has taken two days of the medications prescribed by the ER in Abilene: Levaquin 500 mg daily and Benzonatate 200 mg TID.

Review of Systems:

Constitutional: Denies changes in weight, fatigue, night-sweats.
HEENT: No changes in vision, nasal discharge, headache.
CV: No palpitations, left-sided chest pain/pressure, edema.
Resp: See HPI
GI: No nausea, vomiting, diarrhea, constipation.
GU: No dysuria, increased frequency.
Neuro: No weakness, confusion, numbness, dizziness.
MSK: No weakness, arthralgias, myalgias.
Heme: No easy bruising, easy bleeding.
Skin: No new lesions or rashes.
Endocrine: No polydipsia, polyuria, heat/cold sensitivity.

Social History:

Mr. [redacted] is divorced, lives in [redacted], and has been a truck driver for many years. He lives with his brother and his brother's wife. He has several children, including a daughter who also lives in the area. His family is very supportive and they get along well. As per HPI, he smokes a little less than a pack per day and has for thirty-five years. He does not drink alcohol or use other drugs. He tries to be physically active, and his favorite type of exercise is line dancing. He has "excellent" insurance through his employer.

Family History:

No family history of lung disease or clotting disorders.

The H&P

written by a UNC student and publicly available:

https://www.med.unc.edu/medclerk/education/grading/history-and-physical-examination-h-p-examples/

FHIR systems **SHALL** support, at *minimum*, these eight "Common Clinical Notes":

1. Consultation Note (11488-4)
2. Discharge Summary (18842-5)
3. History & Physical Note (34117-2)
4. Procedures Note (28570-0)
5. Progress Note (11506-3)
6. Imaging Narrative (18748-4)
7. Laboratory Report Narrative (11502-2)
8. Pathology Report Narrative (11526-1)

Assessment & Plan:
Previously healthy 56 year old man presenting with shortness of breath, hemoptysis, and right-sided chest pain, with CT images demonstrating pulmonary emboli and infiltrates.

1. Shortness of breath: The symptoms of dyspnea, hemoptysis, and the right-sided chest pain could all be explained by a pulmonary embolus, and the CT shows evidence that confirms this diagnosis. Management should be to supply supplemental oxygen to maintain adequate oxygen saturation, to control Mr.[redacted]'s chest pain, and to anticoagulate him. Anticoagulation with low molecular weight heparin and Coumadin could be started simultaneously, with both therapies should overlapping for at least five days; we will seek to obtain an INR between 2 and 3. The patient is currently hemodynamically stable, without evidence of right heart strain, so thrombolysis does not seem indicated at this time.
With his history of long drives back and forth to California, Mr. [redacted] is at risk for venous stasis, which is a major source of DVT's; he even presents a history of lower leg pain and swelling. Testing for various causes of hypercoagulability would also be reasonable: these include deficiencies of proteins C and S, antithrombin III, lupus anticoagulant or anticardiolipin antibodies, and factor V mutations, prothrombin mutation, lipoprotein a, and hyperhomocystinemia.
Another important factor to consider would be malignancy, especially as Mr.[redacted] is a long-time smoker. Screening for lung, colon, and prostate cancer would be appropriate.

2. Possible pneumonia: Mr. [redacted] has a history of being febrile, and was diagnosed with pneumonia in Abilene (obtaining these records will be helpful), but does not currently have an elevated white count. The CT demonstrates an infiltrate, possibly consistent with pneumonia. Sputum cultures and blood cultures would be reasonable to obtain, though since he has been taking an antibiotic for presumed community acquired pneumonia, this may be low-yield. His vitals should be monitored regularly. It is possible that all pulmonary symptoms stem from the pulmonary embolus, but it would be reasonable to continue the antibiotic therapy that he has already started in order to cover the possibility of an infectious component.

3. Anemia: Admission labs show hemoglobin at 11.5, with MCV of 89.5, so this anemia is normocytic. The patient denies any history of anemia when questioned.

Notes are a rich source of info

Journal of the American Medical Informatics Association, 26(11), 2019, 1189–1194
doi: 10.1093/jamia/ocz119
Advance Access Publication Date: 12 August 2019
Research and Applications

Research and Applications

Real world evidence in cardiovascular medicine: ensuring data validity in electronic health record-based studies

Tina Hernandez-Boussard,[1,2,3] Keri L Monda,[4,5] Blai Coll Crespo,[4] and Dan Riskin[1,3,6]

Table 1. Cohort identification of diseases and procedures stratified by EHR-S and EHR-U data[a]

Cohort	Occurrence		EHR-S			EHR-U		
	Concept	Patient	Recall (%)	Precision (%)	F1-score (%)	Recall (%)	Precision (%)	F1-score (%)
Hyperlipidemia	2471	837	65.2	99.3	78.7	98.2	99.4	98.8
Hypercholesterolemia	1899	478	55.1	98.0	70.5	90.4	98.8	94.4
Coronary artery disease	1427	465	67.5	99.4	80.4	94.6	96.2	95.4
Diabetes mellitus	4502	1377	80.6	97.9	88.4	97.0	92.6	94.8
Myocardial infarction	523	282	29.8	86.2	44.2	90.4	76.5	82.9
Chronic kidney disease	640	101	40.8	97.6	57.6	92.9	97.9	95.3
Stroke	693	307	36.5	97.2	53.0	95.7	79.6	87.0
Dementia	317	103	62.1	100.0	76.6	93.1	90.0	91.5
Cataract	240	85	28.6	100.0	44.4	96.1	94.9	95.5
CABG[b]	194	73	32.2	100.0	48.7	96.6	95.0	95.8

[a]All comparisons were significant at $P < .0001$.
[b]Coronary artery bypass graft.

A patient presents with shortness of breath…

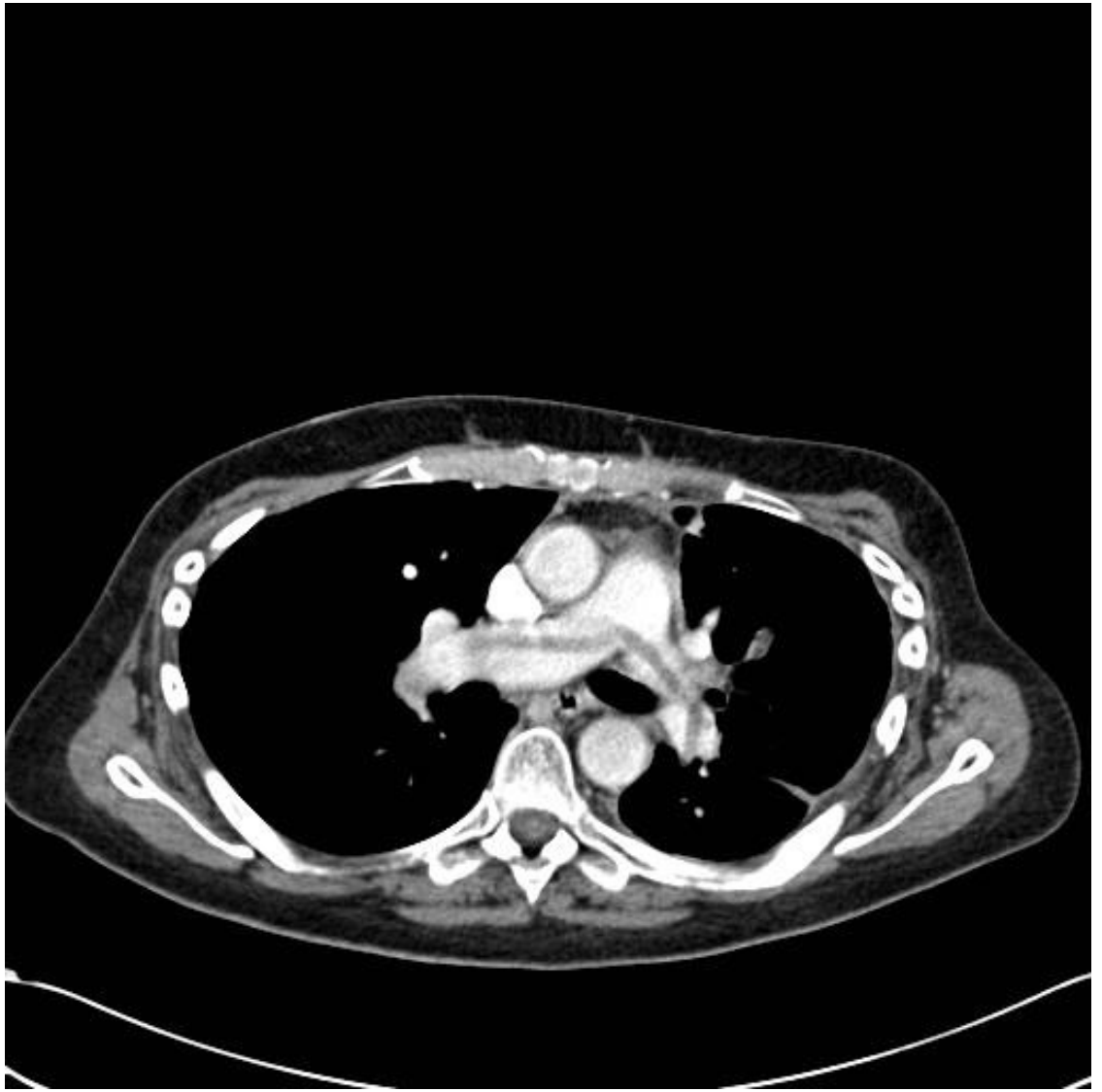
- Presents to urgent care and is transferred to the ED

- Evaluated in the ED
  - History of present illness
  - Review of medical history (including meds and allergies) + social and family hx
  - Physical exam and review of systems
  - Lab results: BMP and CBC
  - Imaging & other studies: CT and EKG

- Admitted to the hospital

CT / Imaging

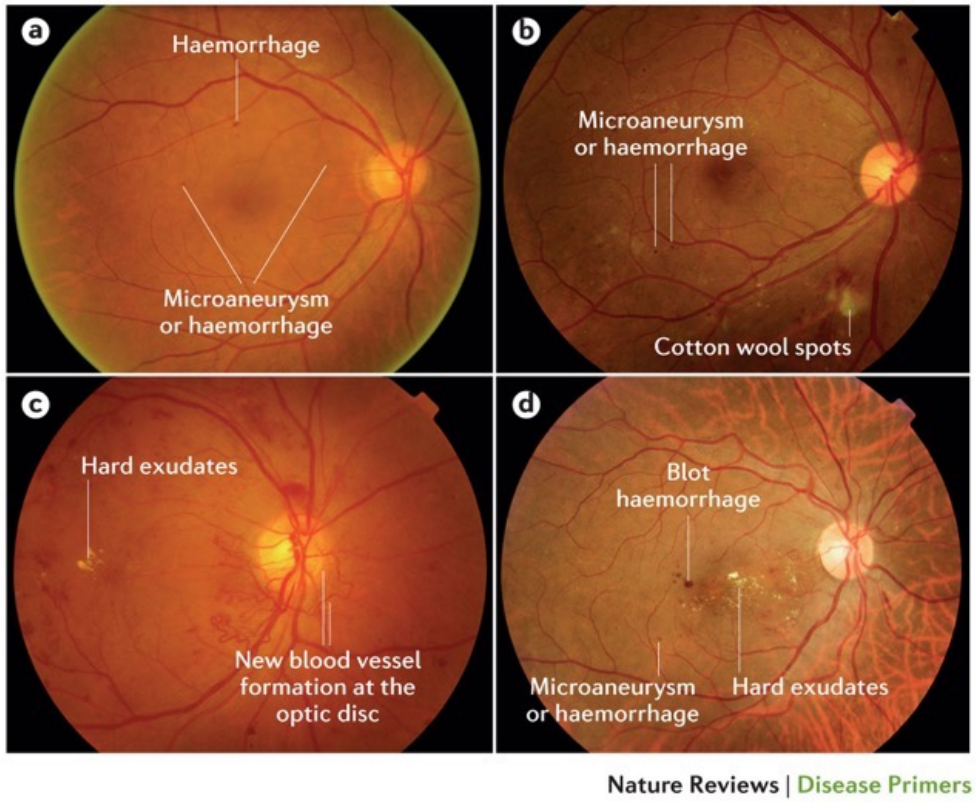
- The scans / images themselves
- Interpretation by radiologist (note)



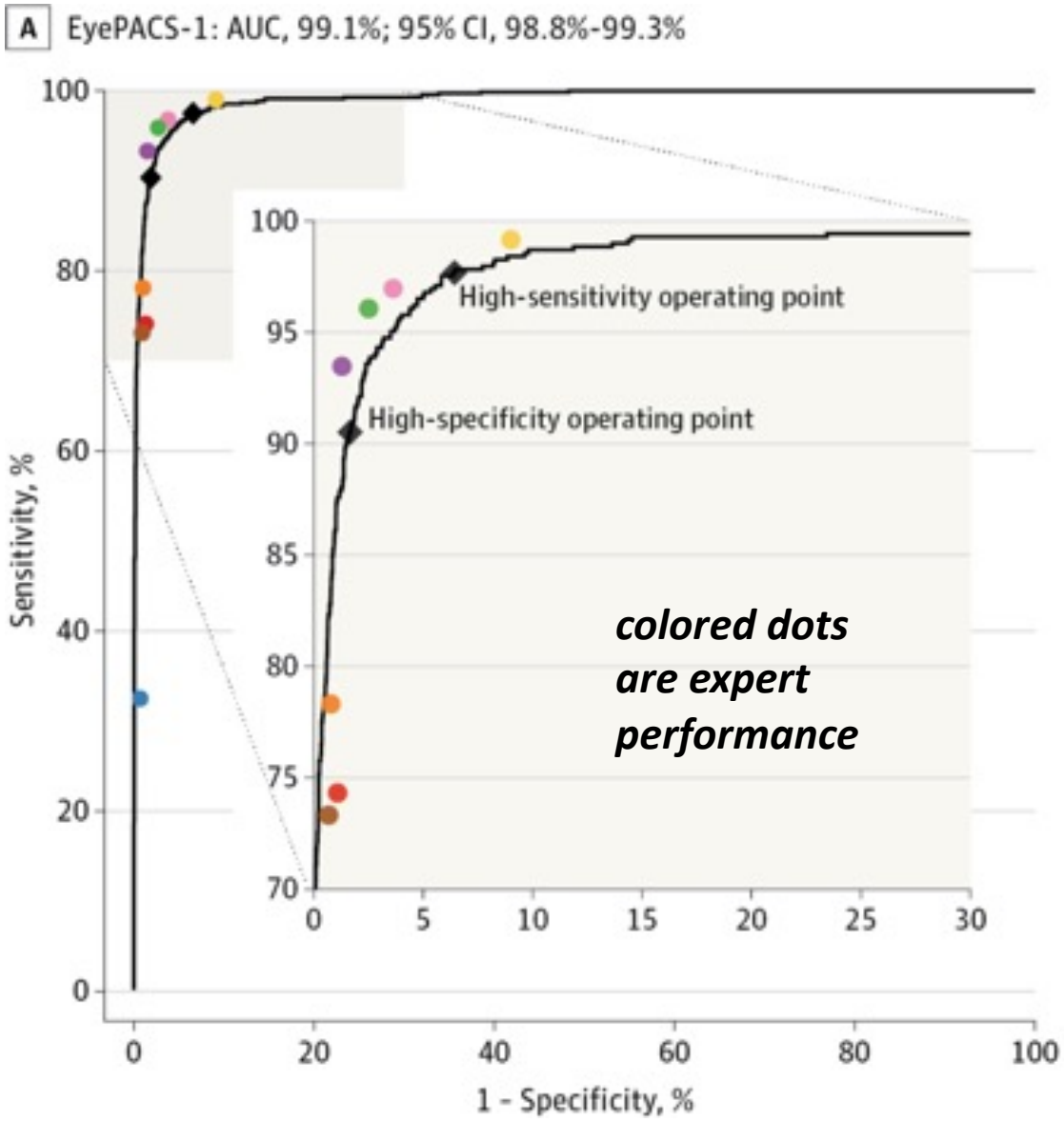
Large non-occlusive pulmonary embolus draped over the bifurcation of the main pulmonary artery; a saddle pulmonary embolus.

https://radiopaedia.org/cases/pulmonary-embolism-saddle-1?lang=us

Retinopathy prediction from fundus images



Wong, T. Y. *et al.* (2016) Diabetic retinopathy
Nat. Rev. Dis. Primers doi:10.1038/nrdp.2016.12



Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA.* 2016;316(22):2402–2410. doi:10.1001/jama.2016.17216

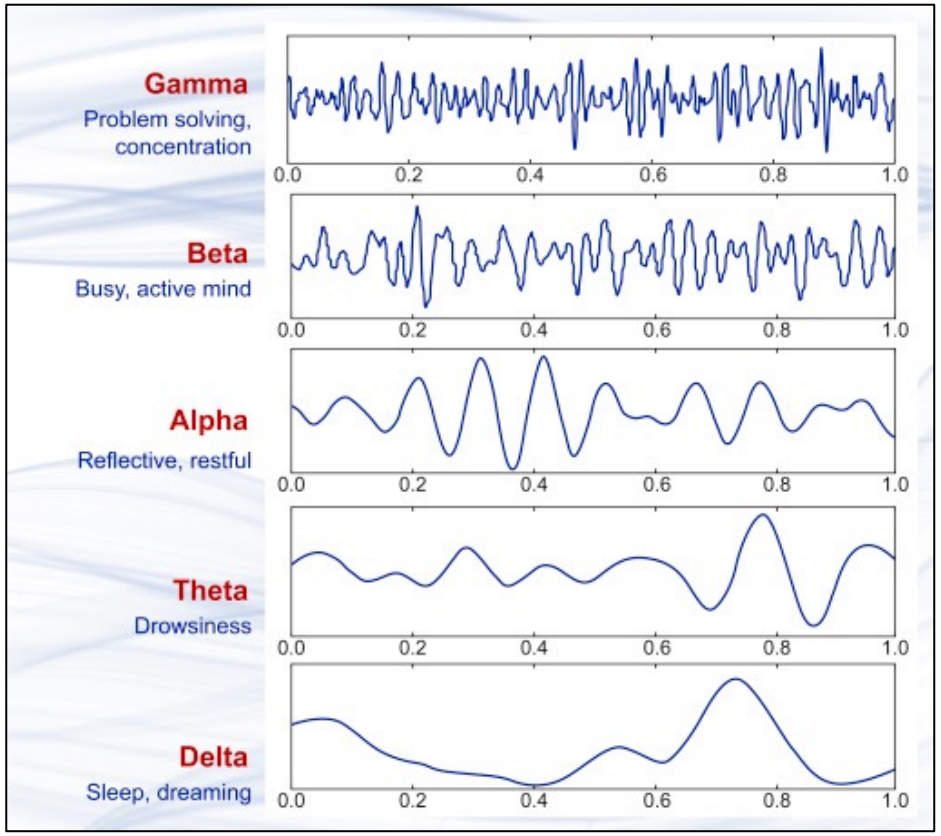
A patient presents with shortness of breath…

- Presents to urgent care and is transferred to the ED

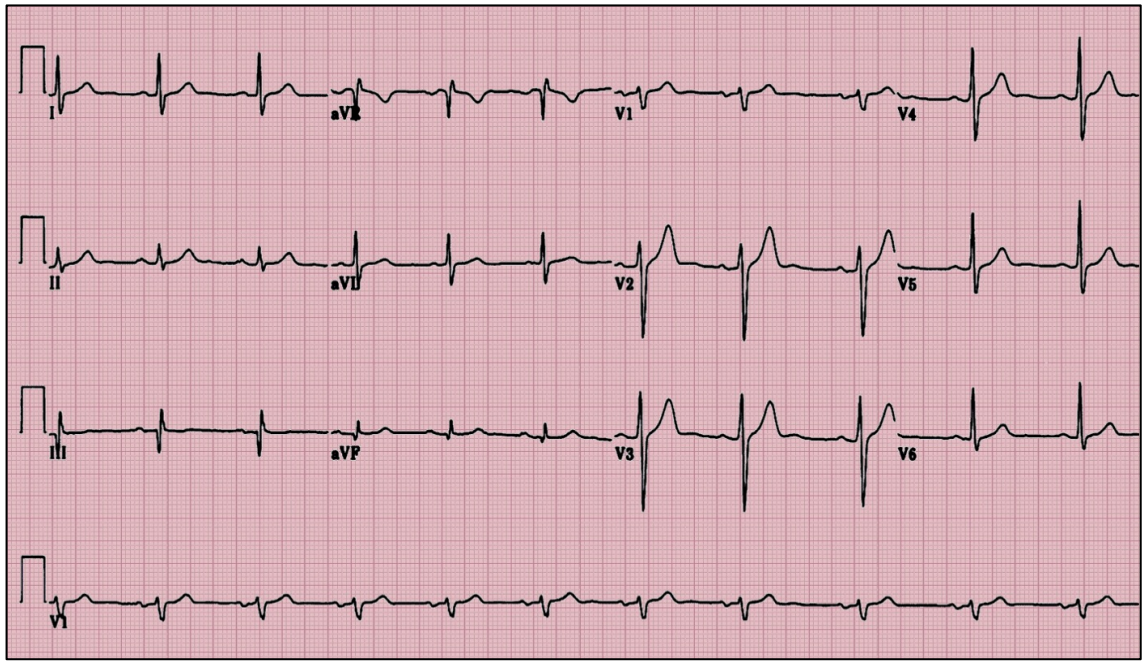
- Evaluated in the ED
  - History of present illness
  - Review of medical history (including meds and allergies) + social and family hx
  - Physical exam and review of systems
  - Lab results: BMP and CBC
  - Imaging & other studies: CT and EKG

- Admitted to the hospital

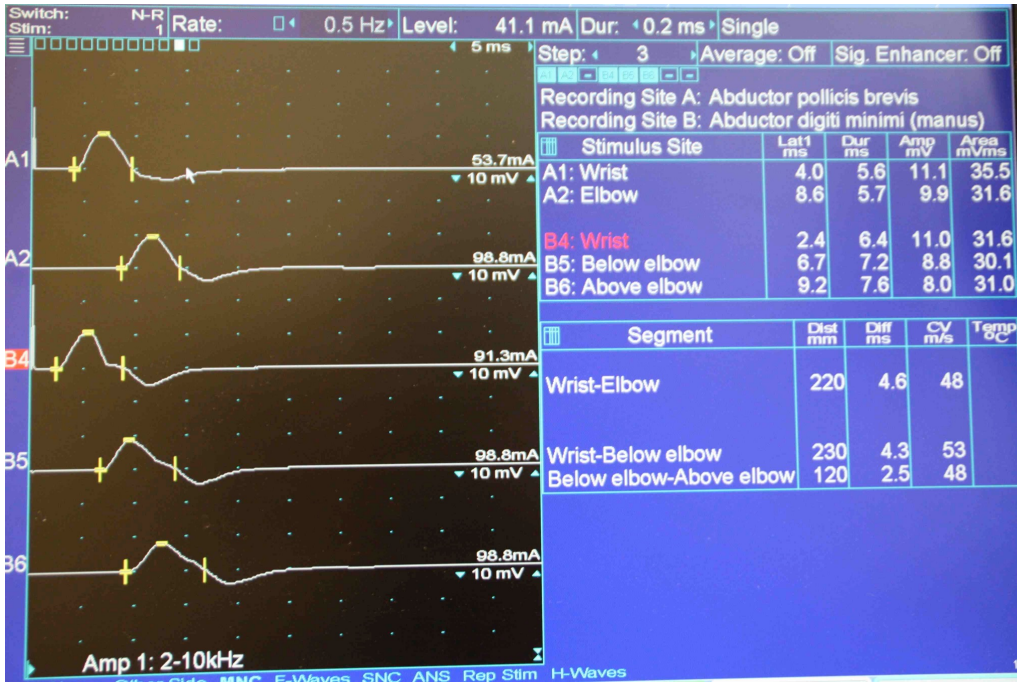
Electrophysiology



EEG



EKG (i.e. ECG)



EMG

Why use unstructured data?

- Reason 1: it contains information not found in structured fields
  - clinial/domain knowledge helps gauge whether/when this info will be useful
  - *example*: use radiological images when predicting COPD prognosis

- Reason 2: it may be available or useful outside of the EHR setting
  - *example*: predict malignancy from images of skin lesions

- Reason 3: it is unaffected (or less affected) by provider decision-making
  - *example*: predict need for surgery from imaging of small bowel obstruction

How do we use unstructured data (in isolation)?

In contrast to structured data…

- Pro: Less manual data wrangling (important caveats apply…)
- Con: Typically larger in size and more difficult to store / load / manage
- Con: Modality-specific knowledge and steps are required to incorporate these data in predictive models

Image Processing: Hierarchical Feature Extraction

pixels

low-level motifs

- edges
- shapes
- textures

high-level motifs

- eyes
- ears
- paws

dog

label

End goal: predict *dog* from *pixels*

Image Processing: Hierarchical Feature Extraction

pixels

low-level motifs

- edges
- shapes
- textures

high-level motifs

- microaneurysm
- exudates
- hemorrhage

retinopathy

label

End goal: predict *retinopathy* from *pixels*

To learn how these neural networks work:

1. Take my course (ML for Health in the CRTP)
2. Sign up for one of the AI Health winter or summer schools
3. For a more rigorous mathematical introduction, take a graduate level course in B&B or another department
4. Many online resources are available

Image Encoder

image

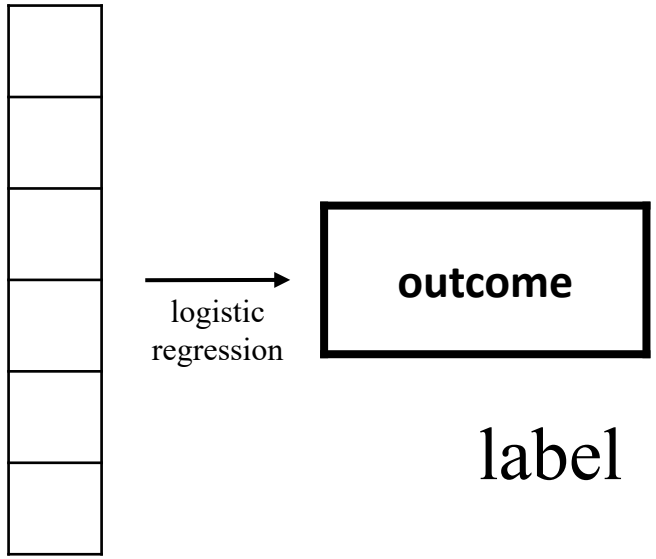
CNN Encoder

vector of features
(indicating high-level motifs)

retinopathy

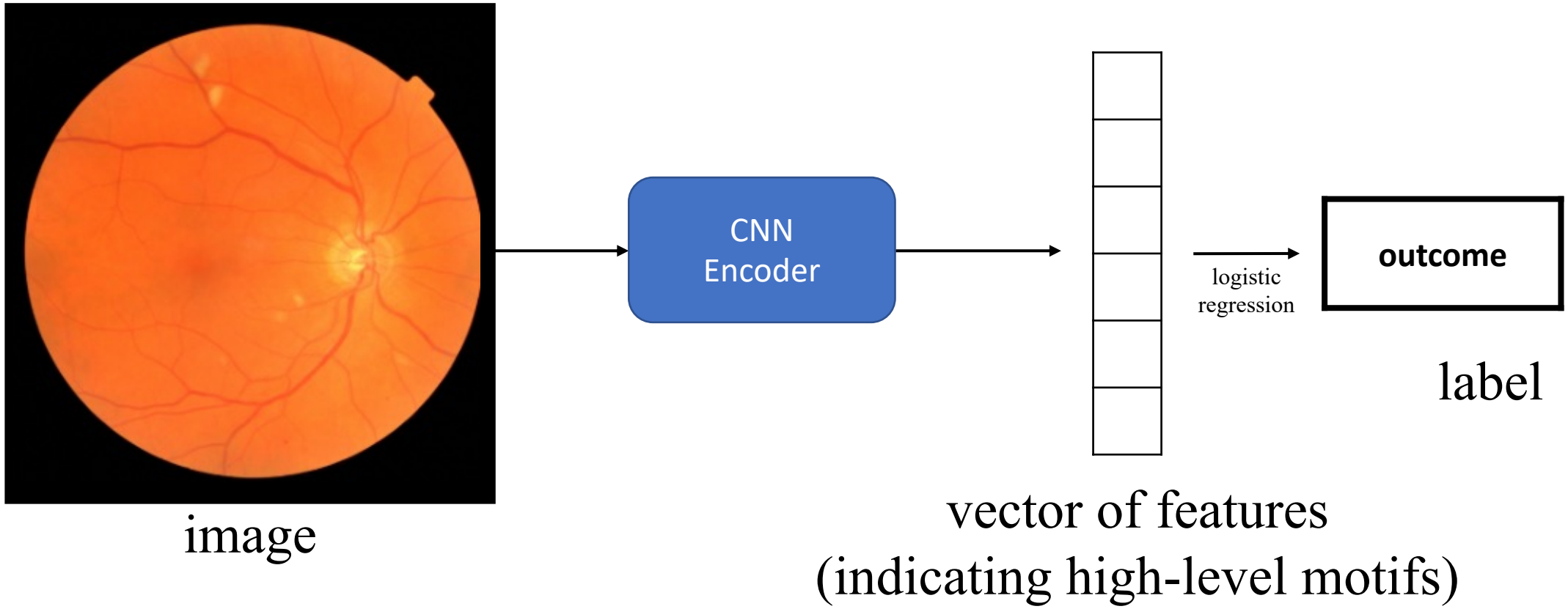
label

A typical prediction model



vector of clinical variables
(e.g. age, sex, BP, HR)

A prediction model for images

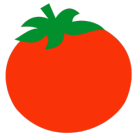


A note on the meaning(s) of “structured”

- Most unstructured data are indexed by spatial location (images), time (time series), or position in a sequence (text)
- We might call these data *dependent*
- As a result, they have *structure* in the sense that measurements that are close together (in space, time, or position) tend to be correlated; and/or there may be characteristic spatial, temporal, or sequential patterns
- Don’t get these two senses of the word “structured” confused!

NLP: Extraction of Word Counts

Dune (2020) Reviews:

- A spicy journey to an amazing future 
- Unlike anything we've seen in science fiction before 
- A smart science fiction film 
- [We've] seen too much of this sort of thing before 
- An effort to create a grand science fiction epic has yielded something cold 

NLP: Extraction of Word Counts

- A spicy journey to an amazing future
- Unlike anything we've seen in science fiction before
- A smart science fiction film
- [We’ve] seen too much of this sort of thing before
- An effort to create a grand science fiction epic has yielded something cold

spicy	journey	amazing	future	unlike	anything	seen	science	fiction	before	smart	film	much	sort	thing	effort	create	grand	epic	yielded	something	cold
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1

A prediction model for documents

Chief Complaint:
Shortness of breath.

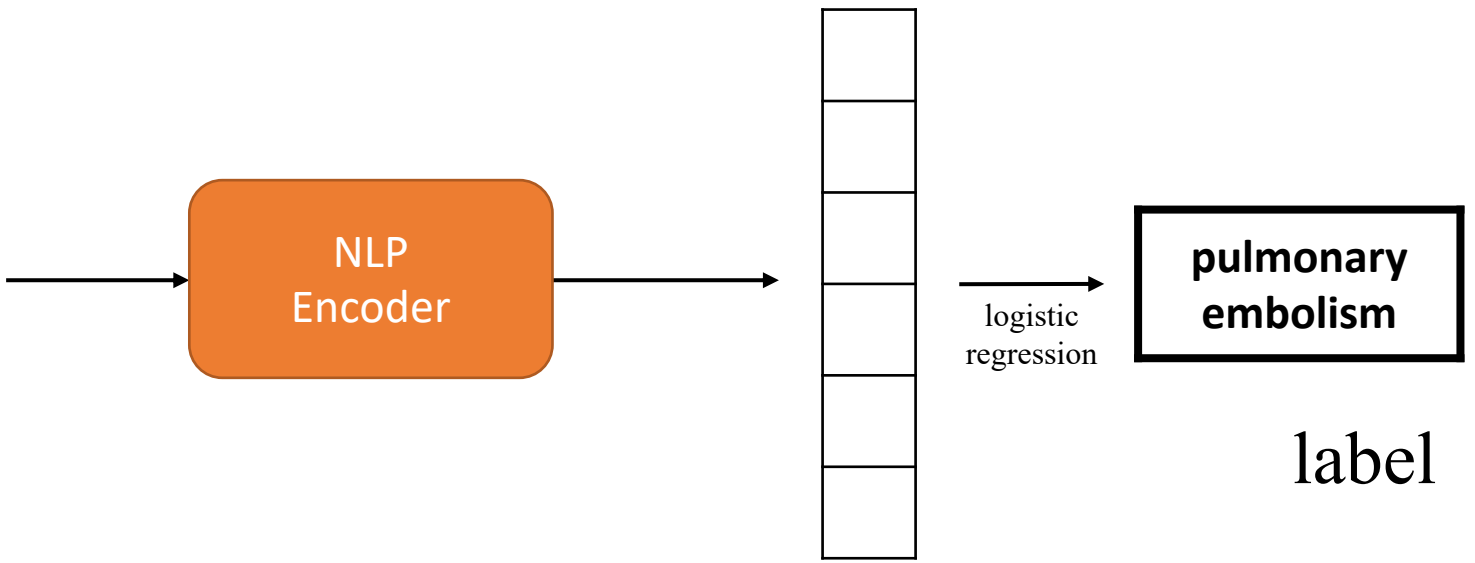
History of the Present Illness:
Mr. [redacted] is a previously healthy 56-year-old gentleman who presents with a four day history of shortness of breath, hemoptysis, and right-sided chest pain. He works as a truck driver, and the symptoms began four days prior to admission, while he was in Jackson, MS. He drove from Jackson to Abilene, TX, the day after the symptoms began, where worsening of his dyspnea and pain prompted him to go to the emergency room. There, he was diagnosed with pneumonia and placed on Levaquin 500 mg daily and Benzonatate 200 mg TID, which he has been taking for two days with only slight improvement. He then drove from Abilene back to Greensboro, where he resides, and continued to experience shortness of breath, right sided chest pain, and hemoptysis. He presented to an urgent care office in town today, and was subsequently transferred to the Moses Cone ER due to the provider's suspicion of PE.

The right-sided pain is located midway down his ribcage, below the axilla. This pain is sharp, about 7/10 in severity, and worsens with movement and cough. Pressing on the chest does not recreate the pain. He feels that the pain has improved somewhat over the past two days. The hemoptysis has been unchanged since it began; there is not frank blood, but his sputum has been consistently blood-tinged. The blood seems redder at night. The dyspnea has been severe, and it is difficult for him to walk more than across a room. He states that he feels as though there is a "rattling" in his chest. At baseline, he experiences no dyspnea on exertion and has no history of COPD or other respiratory problem. He is a smoker, smoking a little less than a pack a day for thirty-five years. Past history is notable for the fact that he experienced transient left lower leg swelling – from below the knee down – and pain several weeks ago during a cross-country haul. He also notes a four day history of decreased appetite, poor sleep, and subjective fever and chills, with a measured fever of 103 in the hospital in Abilene. He had a bout of pneumonia about two months ago, but has been healthy for the most part and denies any chronic medical conditions. Currently he is fairly comfortable, with morphine helping with the pain. He has no history of a clotting disorder, no cardiac history, and denies any chest trauma or aspiration. He has had no sick contacts.

Past Medical History:
1. Hernia repair
2. Bilateral thumb surgeries, secondary to two separate injuries sustained while working with machinery

Medications:
No regular medications, over-the-counter medications, or supplements. Has taken two days of the medications prescribed by the ER in Abilene: Levaquin 500 mg daily and Benzonatate 200 mg TID.

document

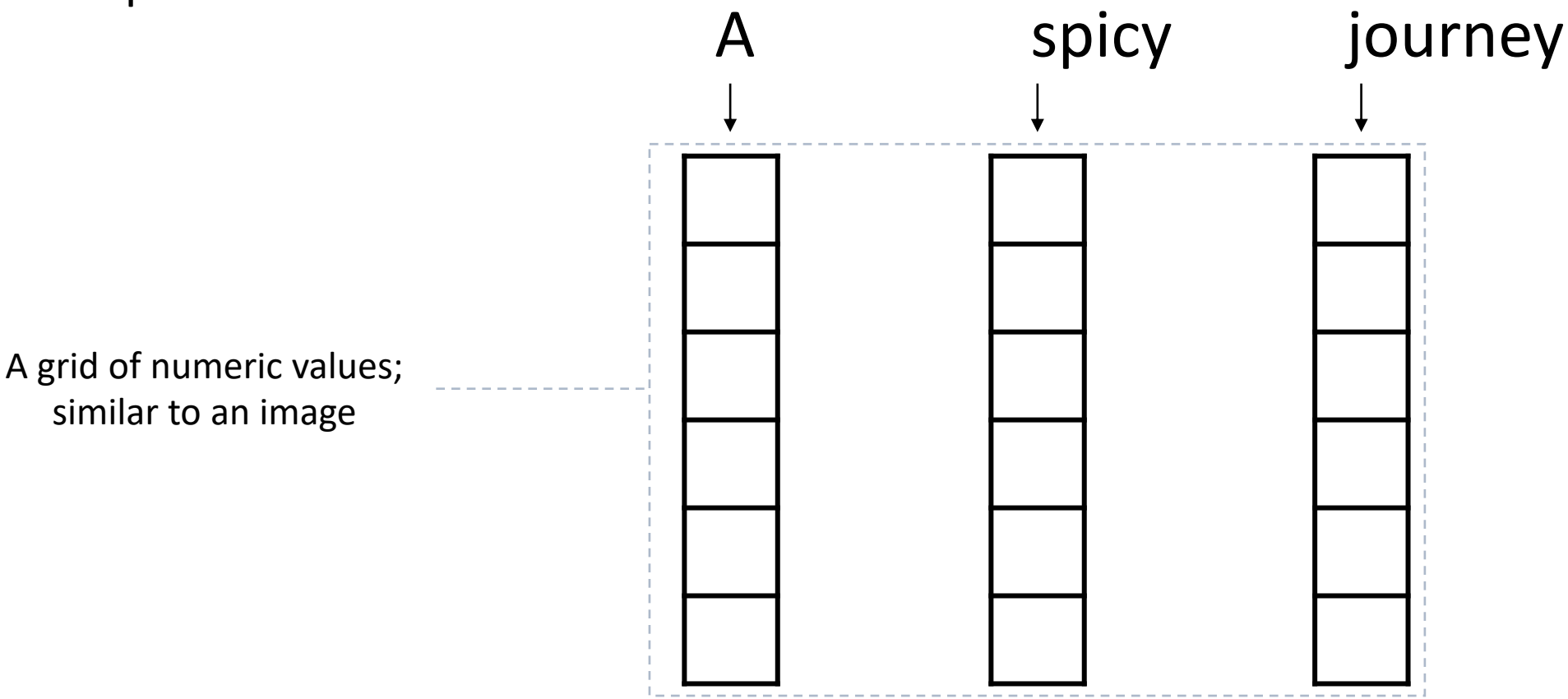


vector of features
(indicating word counts)

label

NLP: Conversion to Word Vectors

- Look up words individually to obtain their vectors
- Construct a sequence of vectors



A prediction model for documents

Chief Complaint:
Shortness of breath.

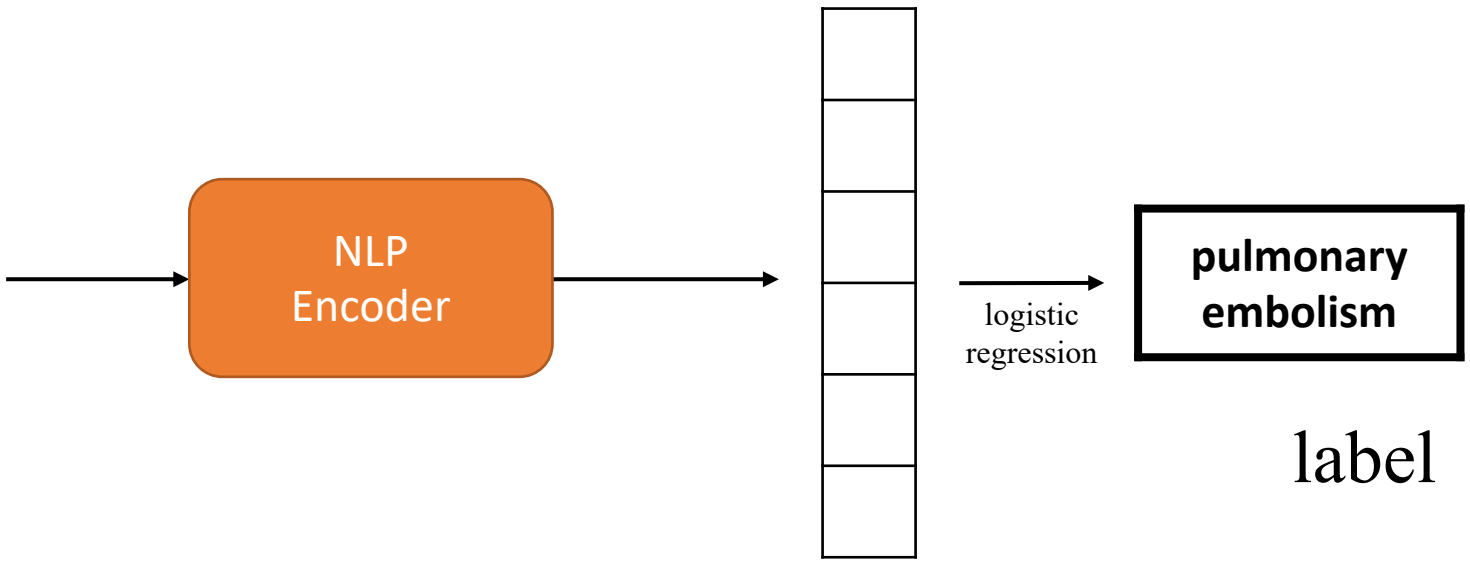
History of the Present Illness:
Mr. [redacted] is a previously healthy 56-year-old gentleman who presents with a four day history of shortness of breath, hemoptysis, and right-sided chest pain. He works as a truck driver, and the symptoms began four days prior to admission, while he was in Jackson, MS. He drove from Jackson to Abilene, TX, the day after the symptoms began, where worsening of his dyspnea and pain prompted him to go to the emergency room. There, he was diagnosed with pneumonia and placed on Levaquin 500 mg daily and Benzonatate 200 mg TID, which he has been taking for two days with only slight improvement. He then drove from Abilene back to Greensboro, where he resides, and continued to experience shortness of breath, right sided chest pain, and hemoptysis. He presented to an urgent care office in town today, and was subsequently transferred to the Moses Cone ER due to the provider's suspicion of PE.

The right-sided pain is located midway down his ribcage, below the axilla. This pain is sharp, about 7/10 in severity, and worsens with movement and cough. Pressing on the chest does not recreate the pain. He feels that the pain has improved somewhat over the past two days. The hemoptysis has been unchanged since it began; there is not frank blood, but his sputum has been consistently blood-tinged. The blood seems redder at night. The dyspnea has been severe, and it is difficult for him to walk more than across a room. He states that he feels as though there is a "rattling" in his chest. At baseline, he experiences no dyspnea on exertion and has no history of COPD or other respiratory problem. He is a smoker, smoking a little less than a pack a day for thirty-five years. Past history is notable for the fact that he experienced transient left lower leg swelling – from below the knee down – and pain several weeks ago during a cross-country haul. He also notes a four day history of decreased appetite, poor sleep, and subjective fever and chills, with a measured fever of 103 in the hospital in Abilene. He had a bout of pneumonia about two months ago, but has been healthy for the most part and denies any chronic medical conditions. Currently he is fairly comfortable, with morphine helping with the pain. He has no history of a clotting disorder, no cardiac history, and denies any chest trauma or aspiration. He has had no sick contacts.

Past Medical History:
1. Hernia repair
2. Bilateral thumb surgeries, secondary to two separate injuries sustained while working with machinery

Medications:
No regular medications, over-the-counter medications, or supplements. Has taken two days of the medications prescribed by the ER in Abilene: Levaquin 500 mg daily and Benzonatate 200 mg TID.

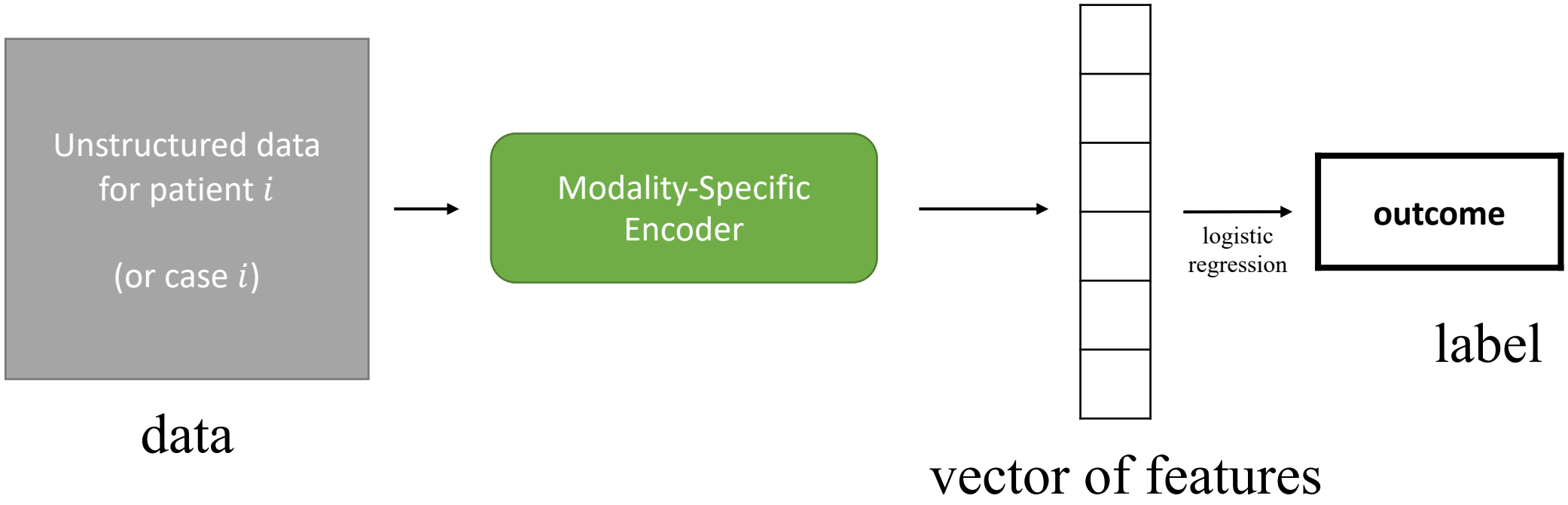
document



vector of features
(indicating document attributes)

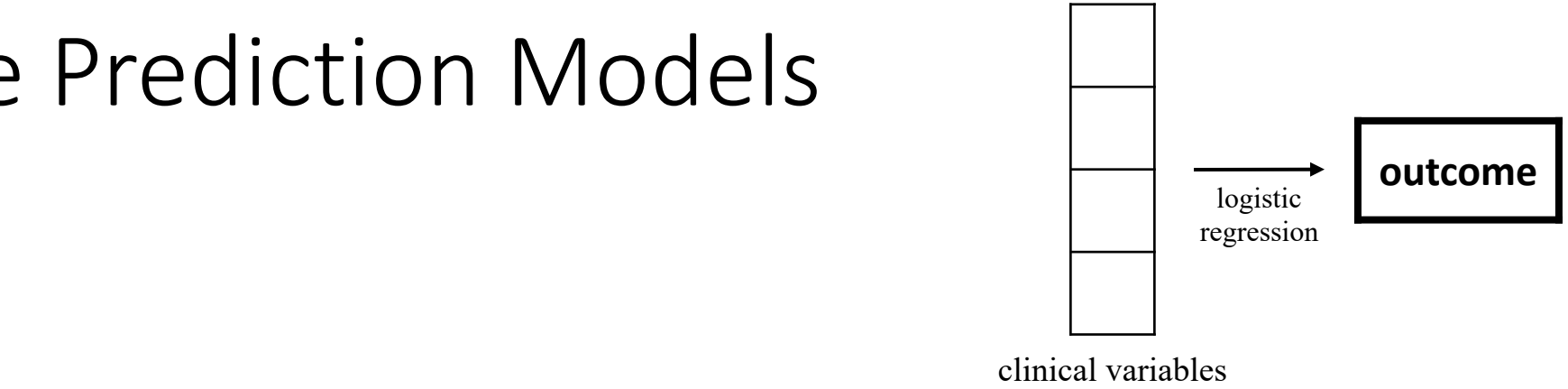
label

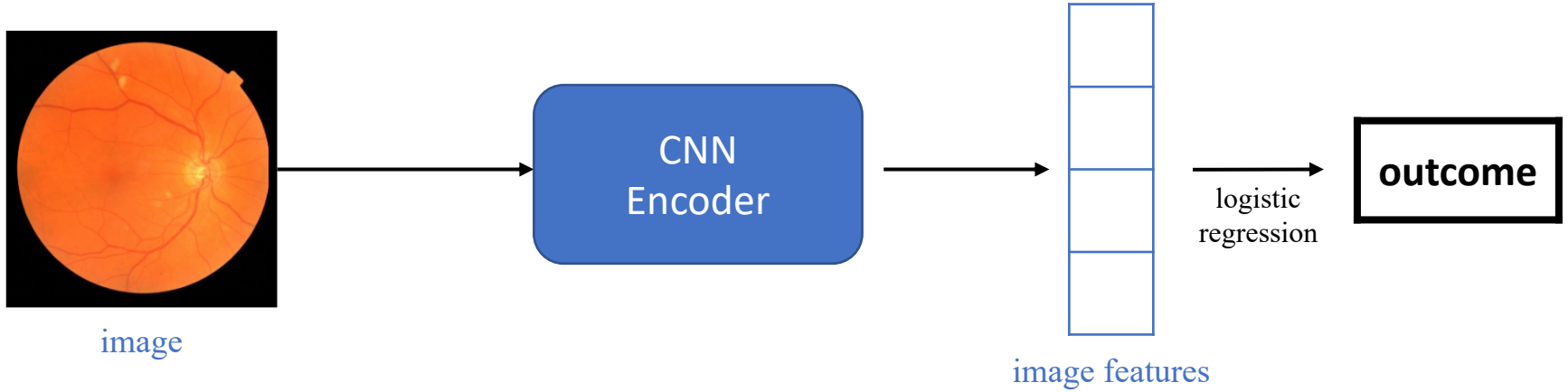
A prediction model for unstructured data

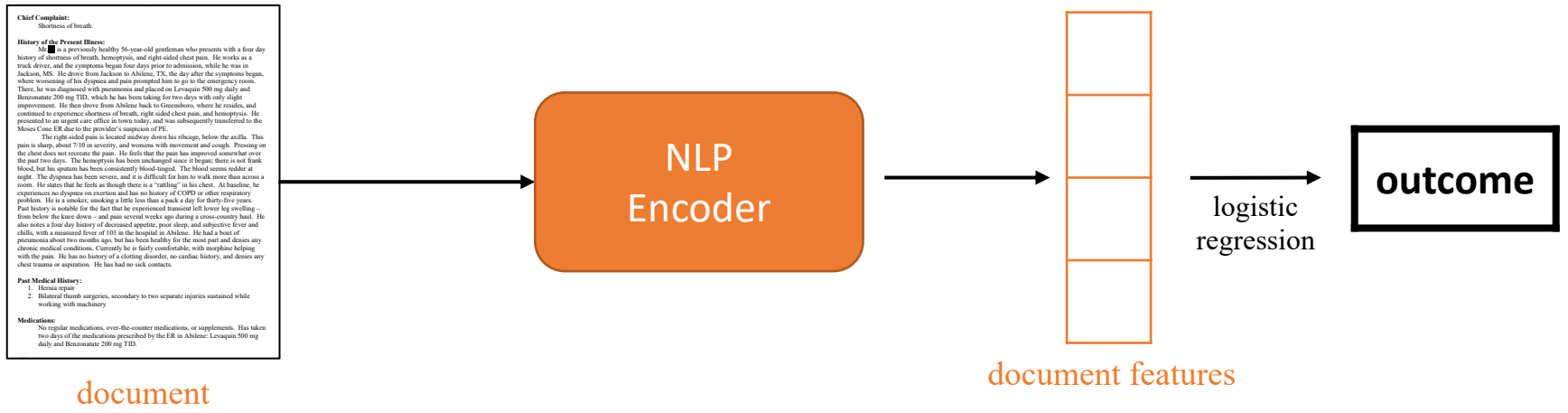


How do we use unstructured data together with structured data?

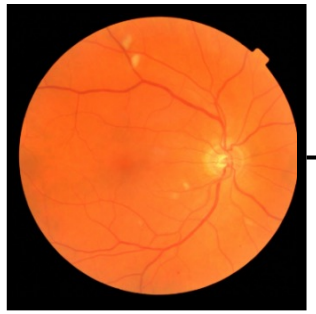
Separate Prediction Models







Multi-Modal Prediction Model



image

Chief Complaint:
Shortness of breath.

History of the Present Illness:
Mr. [redacted] is a previously healthy 56-year-old gentleman who presents with a four day history of shortness of breath, hemoptysis, and right-sided chest pain. He works as a truck driver, and the symptoms began four days prior to admission, while he was in Jackson, MS. He drove from Jackson to Abilene, TX, the day after the symptoms began, where worsening of his dyspnea and pain prompted him to go to the emergency room. There, he was diagnosed with pneumonia and placed on Levaquin 500 mg daily and Benzonatate 200 mg TID, which he has been taking for two days with only slight improvement. He then drove from Abilene back to Greensboro, where he resides, and continued to experience shortness of breath, right sided chest pain, and hemoptysis. He presented to an urgent care office in town today, and was subsequently transferred to the Moses Cone ER due to the provider's suspicion of PE.

The right-sided pain is located midway down his ribcage, below the axilla. This pain is sharp, about 7/10 in severity, and worsens with movement and cough. Pressing on the chest does not recreate the pain. He feels that the pain has improved somewhat over the past two days. The hemoptysis has been unchanged since it began; there is not frank blood, but his sputum has been consistently blood-tinged. The blood seems redder at night. The dyspnea has been severe, and it is difficult for him to walk more than across a room. He states that he feels as though there is a "rattling" in his chest. At baseline, he experiences no dyspnea on exertion and has no history of COPD or other respiratory problem. He is a smoker, smoking a little less than a pack a day for thirty-five years. Past history is notable for the fact that he experienced transient left lower leg swelling – from below the knee down – and pain several weeks ago during a cross-country haul. He also notes a four day history of decreased appetite, poor sleep, and subjective fever and chills, with a measured fever of 103 in the hospital in Abilene. He had a bout of pneumonia about two months ago, but has been healthy for the most part and denies any chronic medical conditions. Currently he is fairly comfortable, with morphine helping with the pain. He has no history of a clotting disorder, no cardiac history, and denies any chest trauma or aspiration. He has had no sick contacts.

Past Medical History:
1. Hernia repair
2. Bilateral thumb surgeries, secondary to two separate injuries sustained while working with machinery

Medications:
No regular medications, over-the-counter medications, or supplements. Has taken two days of the medications prescribed by the ER in Abilene: Levaquin 500 mg daily and Benzonatate 200 mg TID.

document

CNN Encoder

NLP Encoder

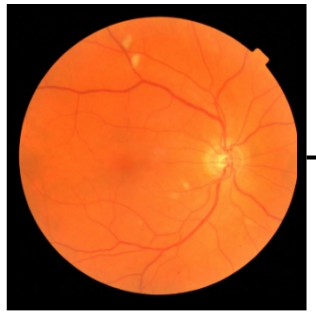
logistic regression

outcome

concatenated, multi-modal features

Multi-Modal Prediction Model

1. Train independently
2. Concatenate
3. Train together



image

Chief Complaint:
Shortness of breath.

History of the Present Illness:
Mr. [redacted] is a previously healthy 56-year-old gentleman who presents with a four day history of shortness of breath, hemoptysis, and right-sided chest pain. He works as a truck driver, and the symptoms began four days prior to admission, while he was in Jackson, MS. He drove from Jackson to Abilene, TX, the day after the symptoms began, where worsening of his dyspnea and pain prompted him to go to the emergency room. There, he was diagnosed with pneumonia and placed on Levaquin 500 mg daily and Benzonatate 200 mg TID, which he has been taking for two days with only slight improvement. He then drove from Abilene back to Greensboro, where he resides, and continued to experience shortness of breath, right sided chest pain, and hemoptysis. He presented to an urgent care office in town today, and was subsequently transferred to the Moses Cone ER due to the provider's suspicion of PE.

The right-sided pain is located midway down his ribcage, below the axilla. This pain is sharp, about 7/10 in severity, and worsens with movement and cough. Pressing on the chest does not recreate the pain. He feels that the pain has improved somewhat over the past two days. The hemoptysis has been unchanged since it began; there is not frank blood, but his sputum has been consistently blood-tinged. The blood seems redder at night. The dyspnea has been severe, and it is difficult for him to walk more than across a room. He states that he feels as though there is a "rattling" in his chest. At baseline, he experiences no dyspnea on exertion and has no history of COPD or other respiratory problem. He is a smoker, smoking a little less than a pack a day for thirty-five years. Past history is notable for the fact that he experienced transient left lower leg swelling – from below the knee down – and pain several weeks ago during a cross-country haul. He also notes a four day history of decreased appetite, poor sleep, and subjective fever and chills, with a measured fever of 103 in the hospital in Abilene. He had a bout of pneumonia about two months ago, but has been healthy for the most part and denies any chronic medical conditions. Currently he is fairly comfortable, with morphine helping with the pain. He has no history of a clotting disorder, no cardiac history, and denies any chest trauma or aspiration. He has had no sick contacts.

Past Medical History:
1. Hernia repair
2. Bilateral thumb surgeries, secondary to two separate injuries sustained while working with machinery

Medications:
No regular medications, over-the-counter medications, or supplements. Has taken two days of the medications prescribed by the ER in Abilene: Levaquin 500 mg daily and Benzonatate 200 mg TID.

document

CNN Encoder

NLP Encoder

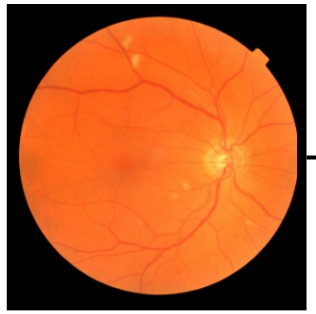
logistic regression

outcome

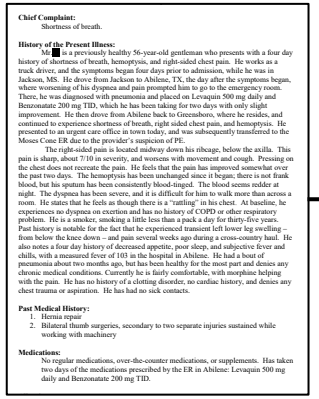
concatenated, multi-modal features

Multi-Modal Prediction Model

1. Train independently
2. Concatenate
3. **Consider adding one more fully-connected, “interaction” layer**
4. Train together
5. **Fine-tune the whole model**



image



Chief Complaint:
Shortness of breath.

History of the Present Illness:
Mr. [redacted] is a previously healthy 56-year-old gentleman who presents with a four day history of shortness of breath, hemoptysis, and right-sided chest pain. He works as a truck driver, and the symptoms began four days prior to admission, while he was in Jackson, MS. He drove from Jackson to Abilene, TX, the day after the symptoms began, where worsening of his dyspnea and pain prompted him to go to the emergency room. There, he was diagnosed with pneumonia and placed on Levaquin 500 mg daily and Benzonatate 200 mg TID, which he has been taking for two days with only slight improvement. He then drove from Abilene back to Greensboro, where he resides, and continued to experience shortness of breath, right sided chest pain, and hemoptysis. He presented to an urgent care office in town today, and was subsequently transferred to the Moses Cone ER due to the provider's suspicion of PE.

The right-sided pain is located midway down his ribcage, below the axilla. This pain is sharp, about 7/10 in severity, and worsens with movement and cough. Pressing on the chest does not recreate the pain. He feels that the pain has improved somewhat over the past two days. The hemoptysis has been unchanged since it began; there is not frank blood, but his sputum has been consistently blood-tinged. The blood seems redder at night. The dyspnea has been severe, and it is difficult for him to walk more than across a room. He states that he feels as though there is a "rattling" in his chest. At baseline, he experiences no dyspnea on exertion and has no history of COPD or other respiratory problem. He is a smoker, smoking a little less than a pack a day for thirty-five years. Past history is notable for the fact that he experienced transient left lower leg swelling – from below the knee down – and pain several weeks ago during a cross-country haul. He also notes a four day history of decreased appetite, poor sleep, and subjective fever and chills, with a measured fever of 103 in the hospital in Abilene. He had a bout of pneumonia about two months ago, but has been healthy for the most part and denies any chronic medical conditions. Currently he is fairly comfortable, with morphine helping with the pain. He has no history of a clotting disorder, no cardiac history, and denies any chest trauma or aspiration. He has had no sick contacts.

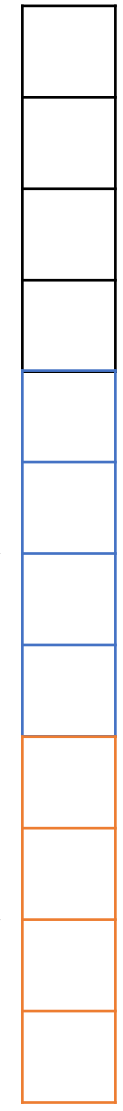
Past Medical History:
1. Hernia repair
2. Bilateral thumb surgeries, secondary to two separate injuries sustained while working with machinery

Medications:
No regular medications, over-the-counter medications, or supplements. Has taken two days of the medications prescribed by the ER in Abilene: Levaquin 500 mg daily and Benzonatate 200 mg TID.

document

CNN Encoder

NLP Encoder

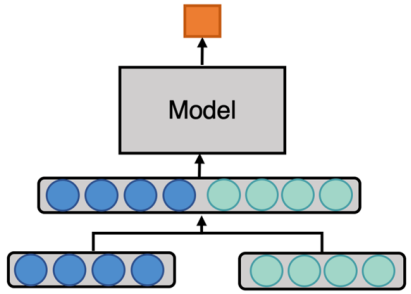


concatenated, multi-modal features

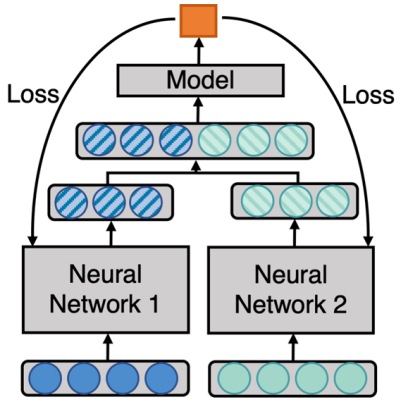
logistic regression

outcome

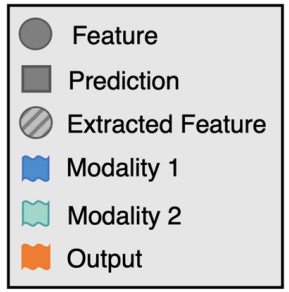
Variations on Multi-Modal Fusion

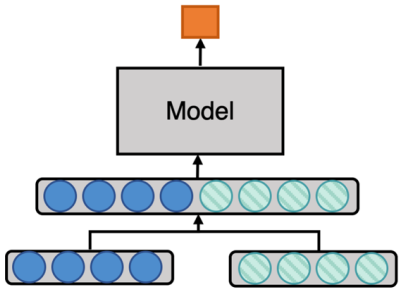


Early Fusion – Type I

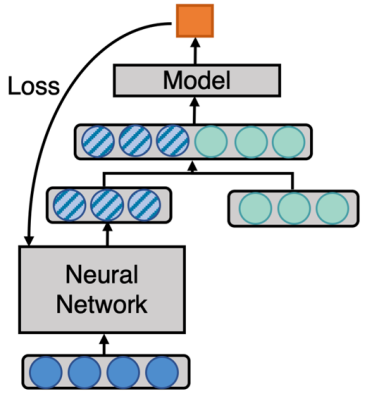


Joint Fusion – Type I

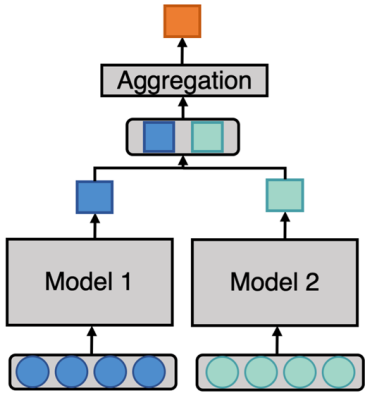




Early Fusion – Type II



Joint Fusion – Type II



Late Fusion

Huang, SC., Pareek, A., Seyyedi, S. *et al.* Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *npj Digit. Med.* **3,** 136 (2020).

Conclusions

- Multi-modal data is common in healthcare
- Unstructured data often contains information not captured in structured fields that is important to a given prediction task
- Unstructured data modalities include images, clinical notes, physiologic time series, and more
- In general, each modality requires an *encoder* specific to that modality to extract features useful for prediction
- To incorporate multiple modalities (e.g. structured data + images + notes) in a single prediction model, we typically concatenate *high-level features* associated with each modality. This is known as *joint fusion*, and it performs much better than the naïve approach of concatenating all of the raw data together in a single vector.

Thank you!

Be in touch -> m.engelhard@duke.edu