DeepPhe:
Cancer Deep Phenotype Extraction from Electronic Medical Records

MPIs:
Harry Hochheiser, PhD, University of Pittsburgh

Jeremy Warner, MD, MS, Lifespan Cancer Institute

Guergana Savova, PhD, Boston Children's Hospital and Harvard Medical School

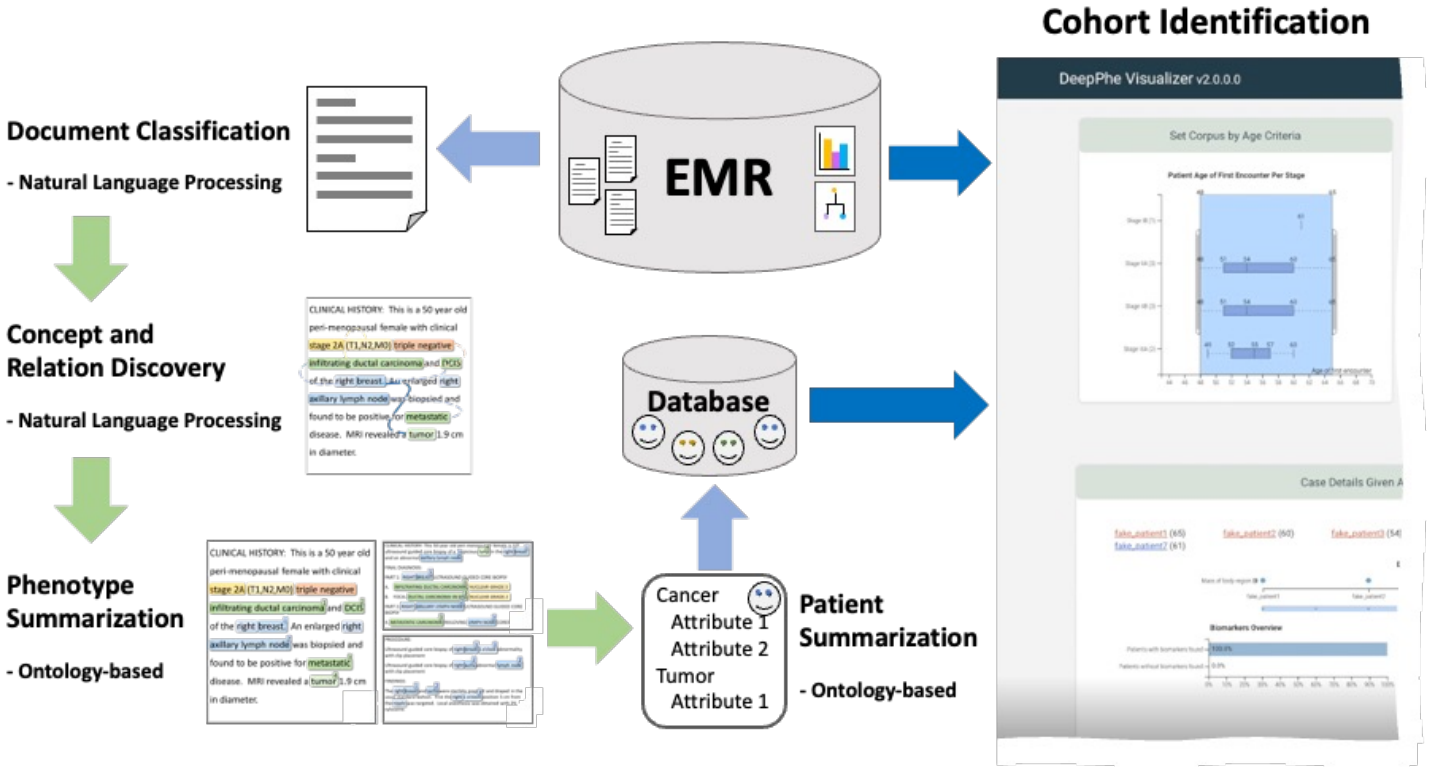
DeepPhe: Cancer Deep Phenotype Extraction from Electronic Medical Records

Rich Data Models

- Natural Language Processing
- Ontology-based Summarization
- Cohort Identification
- Visual Analytics

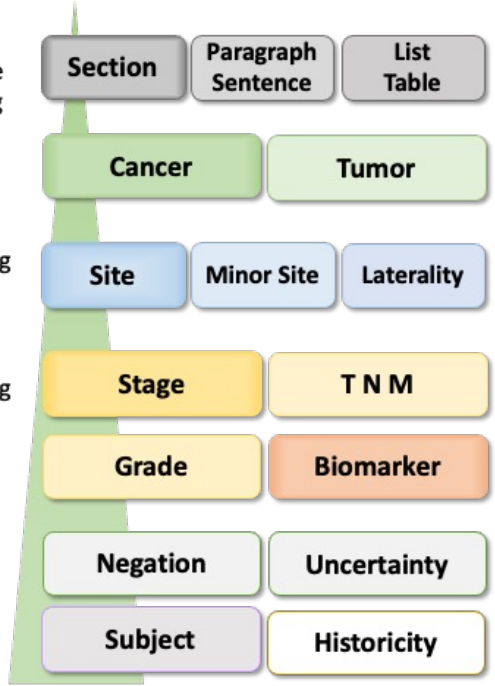
ITCR Grant #U24CA248010

DeepPhe Capabilities



Natural Language Processing (NLP) Pipeline

- Machine Learning Models
- Text Matching
- Pattern Matching
- Rules



CLINICAL HISTORY: This is a 50 year old peri-menopausal female with clinical stage 2A (T1,N2,M0) triple negative infiltrating ductal carcinoma and DCIS of the right breast. An enlarged right axillary lymph node was biopsied and found to be positive for metastatic disease. MRI revealed a tumor 1.9 cm in diameter.

Mock Clinical Note

DeepPhe NLP Extraction: Tumor Characteristics

Attribute	Precision	Recall	F1
Body Site	0.80	0.77	0.79
Laterality	0.81	0.98	0.88
Stage	0.94	0.78	0.85
T-stage	0.96	0.90	0.93
N-stage	0.93	0.92	0.93
M-stage	0.94	0.90	0.92

F1 Target: 0.75

TNM: tumor, (lymph) nodes, metastases

DeepPhe NLP Extraction: Cancer Characteristics

Attribute	Precision	Recall	F1
Body Site	0.87	0.61	0.72
Laterality	0.76	0.98	0.93
Diagnosis	0.90	0.96	0.93
Quadrant	0.97	0.97	0.97
Clockface	0.89	0.97	0.93
ER	0.98	0.98	0.98
PR	0.95	0.98	0.97
HER2	0.87	0.93	0.90

F1 Target: 0.75

ER: estrogen receptor
PR: progesterone receptor
HER2: epidermal growth factor receptor 2 amplification

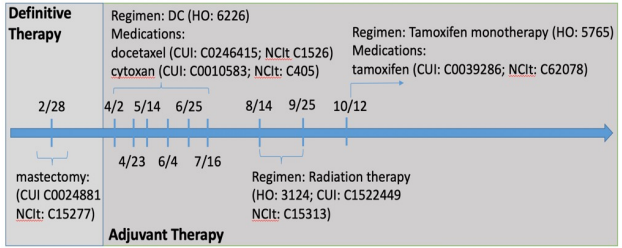
DeepPhe NLP Extraction: Medications

Attribute	Precision	Recall	F1
Drug	0.86	0.88	0.87
Dosage	0.95	0.79	0.86
Duration	0.6	1.0	0.75
Form	0.88	0.86	0.87
Frequency	0.82	0.83	0.83
Route	0.74	0.76	0.75
Strength	0.91	0.96	0.94

F1 Target: 0.75

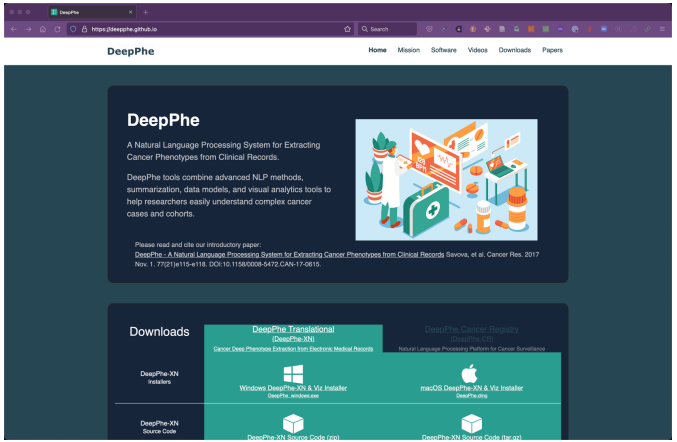
Other Functionalities

- Episode classification
- Tumor marker extraction
- Radiotherapy extraction
- Comorbidities extraction
  - through cTAKES, https://ctakes.apache.org/
- Temporality
  - rough temporality through cTAKES
  - treatment timelines (2023-2024)



- Clinical genomics extraction and normalization (2023)

Using DeepPhe



- deepphe.github.io
  - wizard installer (demo)
- Users/collaborators wanted
  - please get in touch at harryh@pitt.edu
- Feedback
  - https://deepphe.github.io/survey/

Translational Studies

- Can we use DeepPhe to extract data relevant to key translational science questions in cancer care?
  - Breast and Ovarian Cancer
  - Melanoma
- Unstructured + structured EMR data

Breast (BrCa) and Ovarian Cancer (OvCa)

- Frequency/timing of
  - Tumor marker and clinical genomic assessment
  - Associations with comorbidities and treatment

- Example: Tew, et al. **PARP Inhibitors in the Management of Ovarian Cancer: ASCO Guideline (2020)**

  - Patients with ovarian cancer and pathogenic *BRCA1/BRCA2* mutations should be treated with olaparib

Melanoma

- Characterize alignment of treatment and mutations
  - targeted, immunotherapy, both, in what order
  - *BRAF*, *NRAS*, *KIT*, etc.

- Comparison
  - comorbidities
  - other clinical characteristics

Translational Science Data

- Selection:
  - ICD-9-CM/10-CM for diagnosis in absence of tumor registry
  - Tumor registry: ICD-O-3 for diagnosis

- Breast/Ovarian Cancer diagnosed between 2000-2020
  - UPMC – 54,846 pts w/ BrCa, 8,875 pts w/ OvCa
  - VUMC, University of Minnesota access requests underway

- Melanoma diagnosed between 2010-2020
  - UPMC – 16,293 pts
  - VUMC:
    - preliminary analysis of 5,840 cases
    - complete request underway
  - University of Minnesota TBD

Set-aside Project: Reportable Cancer and Recurrence Detection

- PI: John D. Osborne PhD, University of Alabama Birmingham
- Assess utility of DeepPhe descriptions to detect reportable cancers
- Implement and evaluate cancer recurrence detection algorithm
  - to be added to DeepPhe

Related Work: DeepPhe-CR (DeepPhe for Cancer Registries)

- Containerized DeepPhe web services via REST API
- Support registry abstraction efforts
- Collaboration with SEER central cancer registries in
  - Kentucky
  - Louisiana
  - Massachusetts

https://deepphe.github.io/services/deepphe-cr/

Grant #UH3CA243120

Team

Boston Children's Hospital	Guergana Savova (MPI), Sean Finan, Timothy Miller, David Harris, Eli Goldner, Dennis Johns
Dana-Farber/MGB	Elizabeth Buchbinder, Danielle Bitterman
Lifespan/Brown	Jeremy Warner (MPI), Don Dizon, James Yu, Sanjay Mishra, Sandeep Jain
University of Pittsburgh	Harry Hochheiser (MPI), John Levander,
University of Minnesota	Piet de Groen
Vanderbilt University	Douglas Johnson, Alicia Beeghly
University of Alabama	John Osborne, Christopher Coffee, Gaurav Goyal, Sarah Frittro

Guergana.Savova@childrens.harvard.edu
Harryh@pitt.edu
Jeremy_warner@brown.edu