Artificial Intelligence for Near-Real Time Cancer Surveillance: Challenges and Opportunities

Georgia Tourassi, PhD

Director, Health Data Sciences Institute

Oak Ridge National Laboratory

Presented at the NYSDS Summit

June 13, 2019

Outline

- *DOE-NCI partnership*
- *National Cancer Surveillance Program*
- *Research Highlight: AI solutions to support cancer surveillance*
- *Final thoughts and next directions*

Health Data Sciences Institute
Oak Ridge National Laboratory

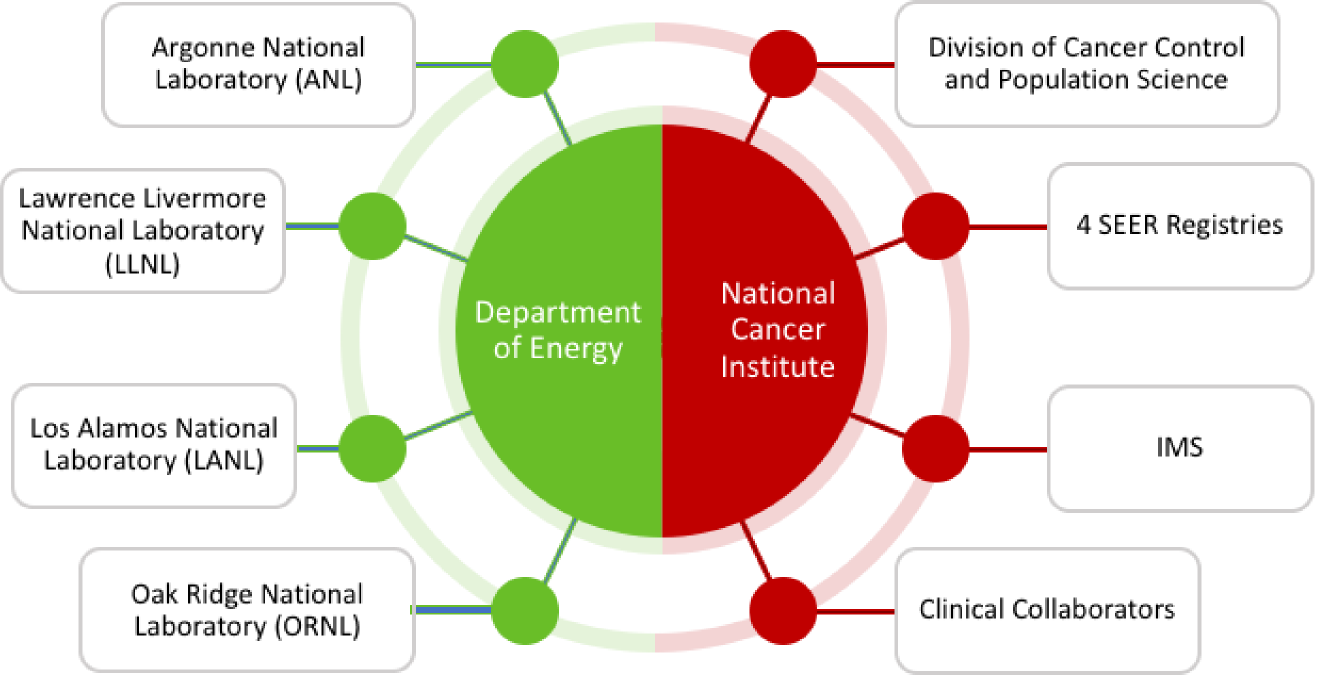
DOE-NCI Partnership: Enable the most challenging deep learning problems in cancer research to run on the most capable supercomputers in the DOE

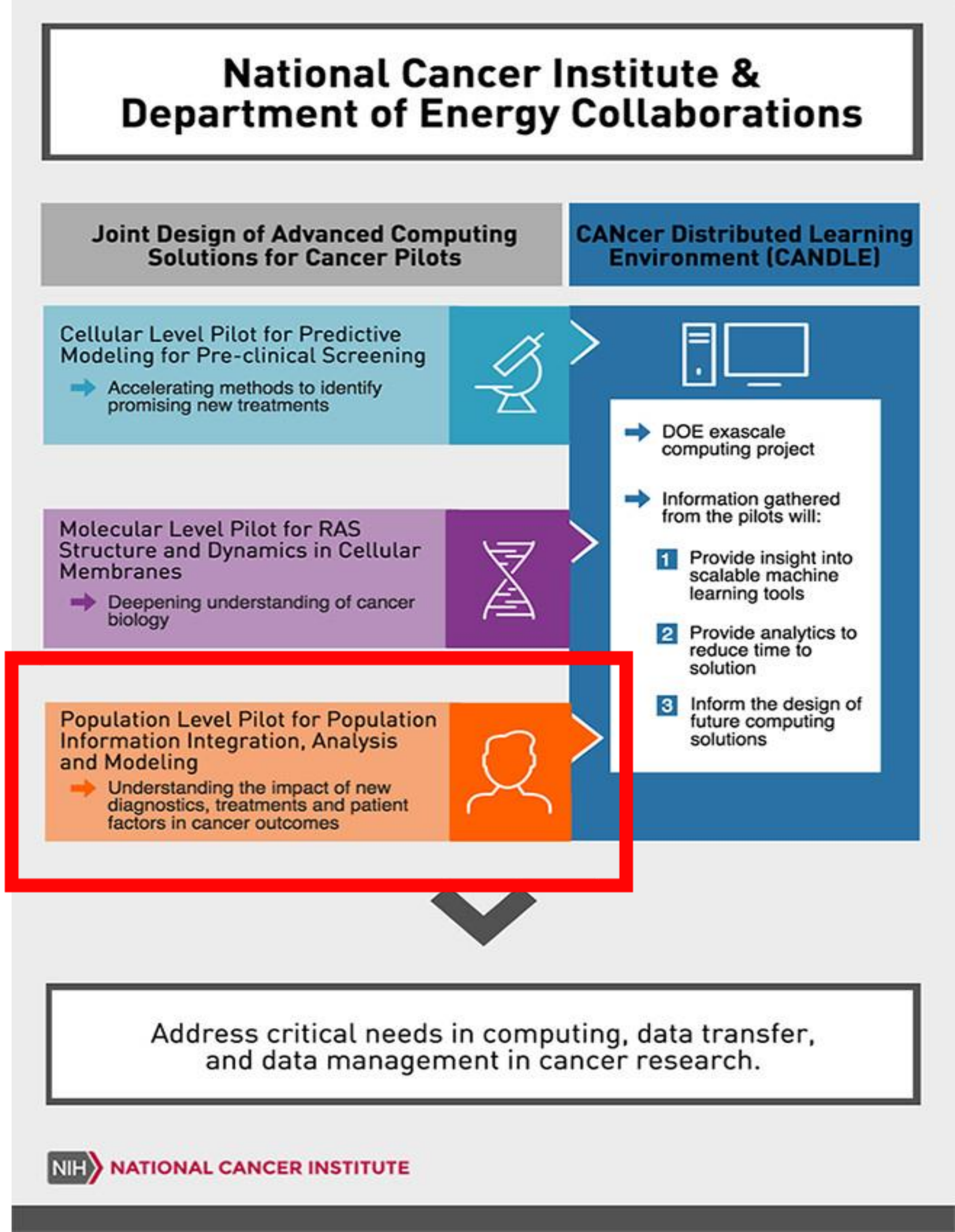
Computing advances for solving the most pressing health issues of our Nation

DOE-NCI partnership

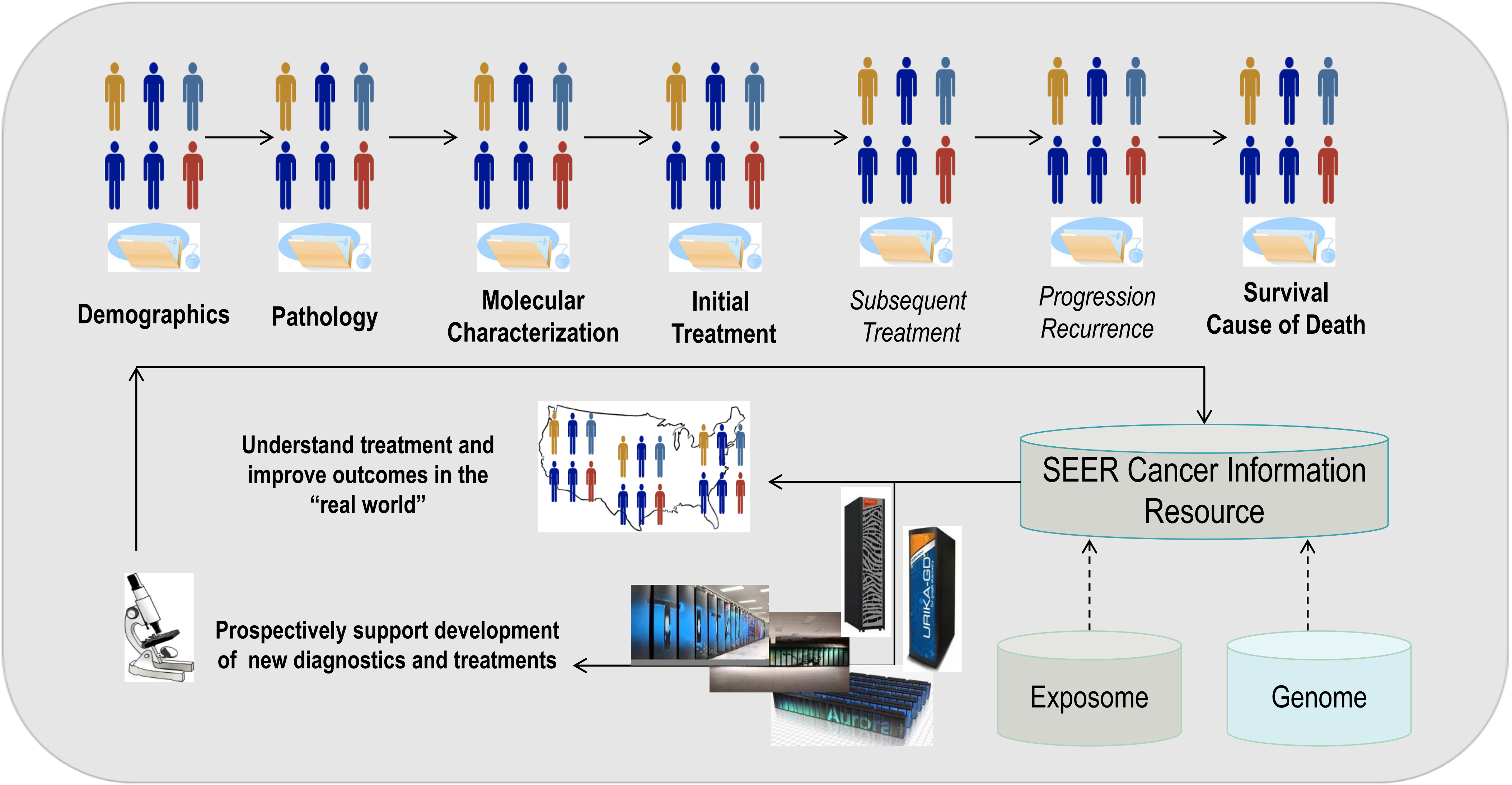
Health and healthcare delivery challenges as a driver for US computing leadership

DOE-NCI partnering entities





AI to support national cancer surveillance



Improve the effectiveness of cancer treatment in the "real world" through computing

Overarching Goal

- **Short-Term:**
  - Deliver a scalable AI solution for large scale, near-real time information capture from unstructured clinical text with state-of-the-art clinical accuracy to semi-automate the cancer surveillance program
- **Long-Term:**
  - Scalable and precise phenotype information extraction to understand the effects of genetic and epigenetic changes on tumor behavior and responsiveness.

- **Critical Challenge:**
  - How to scale across
    - volumes and types of text documents,
    - information extraction tasks / phenotypes
    - cancer registries

State-of-the-Art Approaches in Clinical NLP

- **Current NLP thinking is TASK-specific**
- **Rule-based - **effective but require intense domain expert involvement
  - Task-specific dictionaries of phrases and medical terms
  - *Manual effort not easily scalable across tasks*
- **Traditional machine learning** - scalable but require intense feature engineering
  - N-gram based
  - Concept-extraction-based methods
- **Deep Learning -** scalable with enough compute power and ***enough** data*
  - Does not require dictionaries, not susceptible to misspellings etc.
  - Lots of new DL architectures proposed for NLP
  - No clear winner – depends on the global semantics required for the task at hand

Path NLP is an outstanding challenge

J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Technical Note

The feasibility of using natural language processing to extract clinical information from breast pathology reports

Julliette M. Buckley, Suzanne B. Coopey, John Sharko, Fernanda Polubriaginof, Brian Drohan, Ahmet K. Belli, Elizabeth M. H. Kim, Judy E. Garber[1], Barbara L. Smith, Michele A. Gadd, Michelle C. Specht, Constance A. Roche, Thomas M. Gudewicz[2], Kevin S. Hughes

Departments of Surgical Oncology and [2]Surgical Pathology, Massachusetts General Hospital, [1]Department of Surgical Oncology, Dana Farber Cancer Institute, Boston, Massachusetts, USA

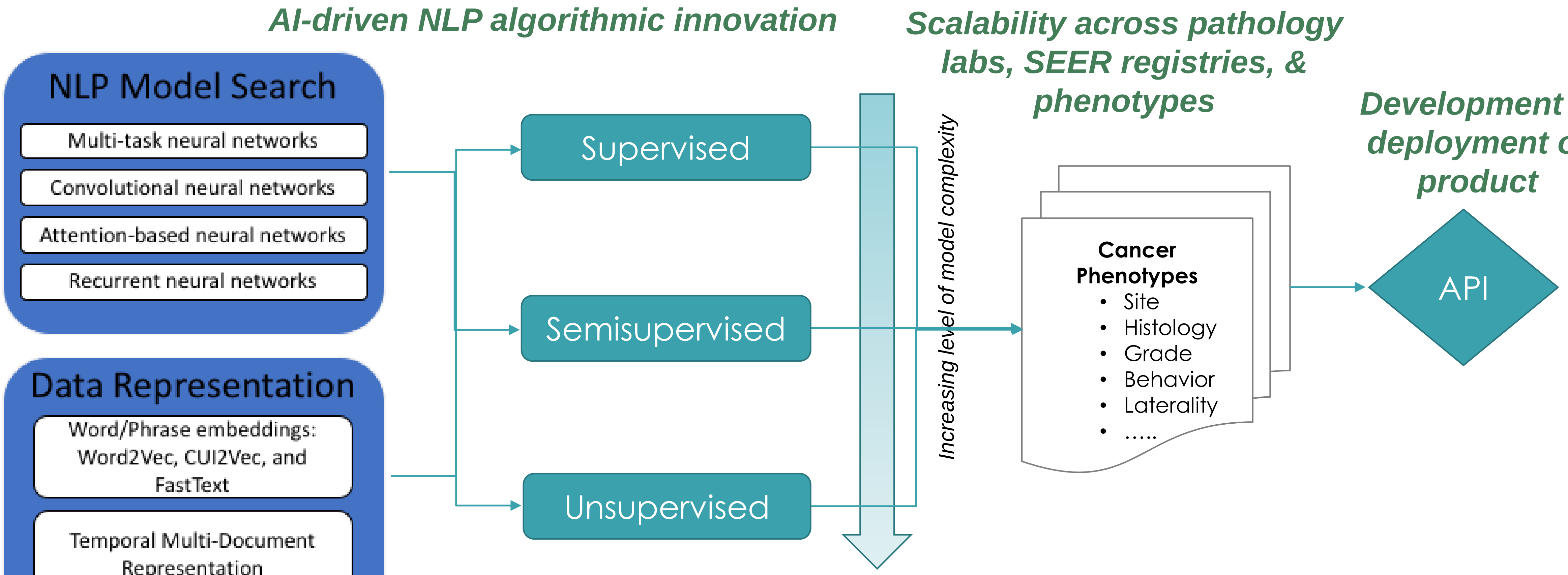
E-mail: *Kevin S. Hughes - kshughes@partners.org
*Corresponding author

In 76,333 breast pathology reports, multiple entities were identified that represented each of the significant buckets. Excluding typographical errors and spacing errors, we identified 124 ways of saying invasive ductal cancer; 95 ways of saying invasive lobular cancer; 52 ways of saying DCIS; 14 ways of saying severe ADH; 53 ways of saying lobular carcinoma *in situ*; 17 ways of saying atypical lobular hyperplasia and 14 ways of saying atypical ductal hyperplasia [Table 1]. Examples of ways to describe ADH and invasive carcinoma are shown in Tables 2 and 3.

In addition, we identified 21 ways of negating a diagnosis when the words appeared before the diagnosis (e.g., No evidence of invasive ductal carcinoma), and an additional 12 ways of negating the diagnosis when the words fell after the diagnosis (e.g., ADH was not seen). As each entity can potentially be negated by a pre- or postnegative one, must multiply the number of ways of stating the negation by the number of ways of describing that particular diagnostic entity. For example, with invasive ductal cancer; that means 124 ways of saying IDC multiplied by 33 ways of saying "not" gives a total of 4092 potential ways to say IDC was not present.

Cancer surveillance programs deals with 70+ cancer sites and 500+ histologies!

AI and HPC for clinical text understanding @ scale



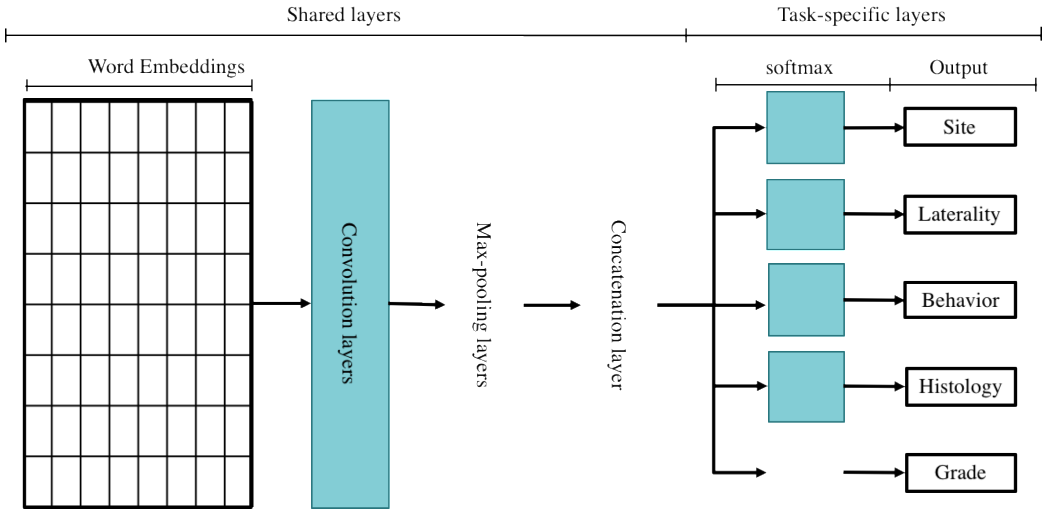
70 cancer sites (306 subsites); 515 histologies; 9 grades; 7 lateralities; 4 behaviors

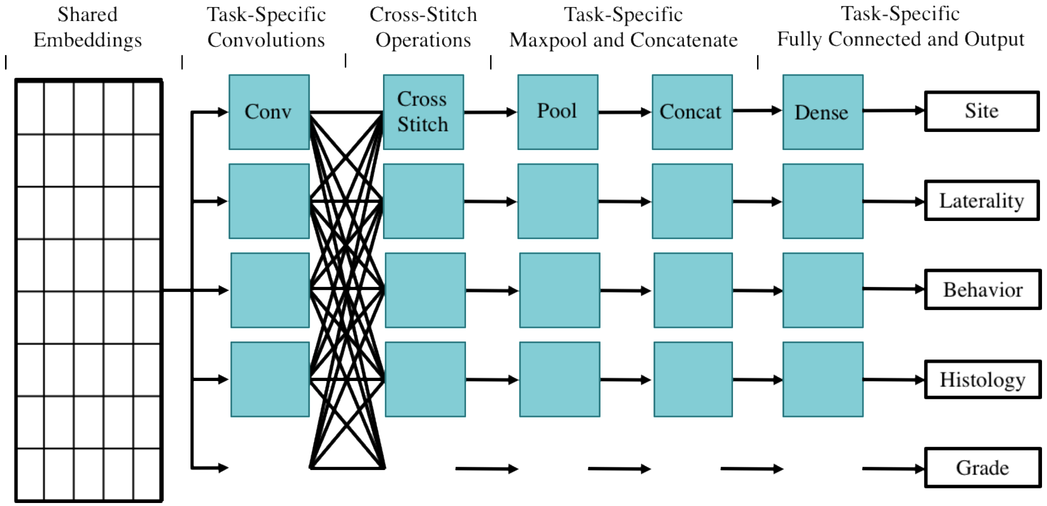
>10,000 cancer phenotypes observed in based only on 5 attributes

Extension to other NLP tasks to extract more data elements (e.g., biomarkers) will increase the number and complexity of cancer phenotypes observed – **combinatorial explosion in computational cancer phenotyping**

Methodology

- Two different implementations of a **multi-task** convolutional neural network
  - Hard-Parameter Sharing
  - Cross-stitch
- Minimal pre-processing
- Simultaneous learning of 5 information extraction tasks:
  - site, histology, behavior, laterality, grade
- Gold standard: The variables coded in the registry abstract
- Benchmarking against traditional machine learning algorithms
- Testing within and across SEER registries



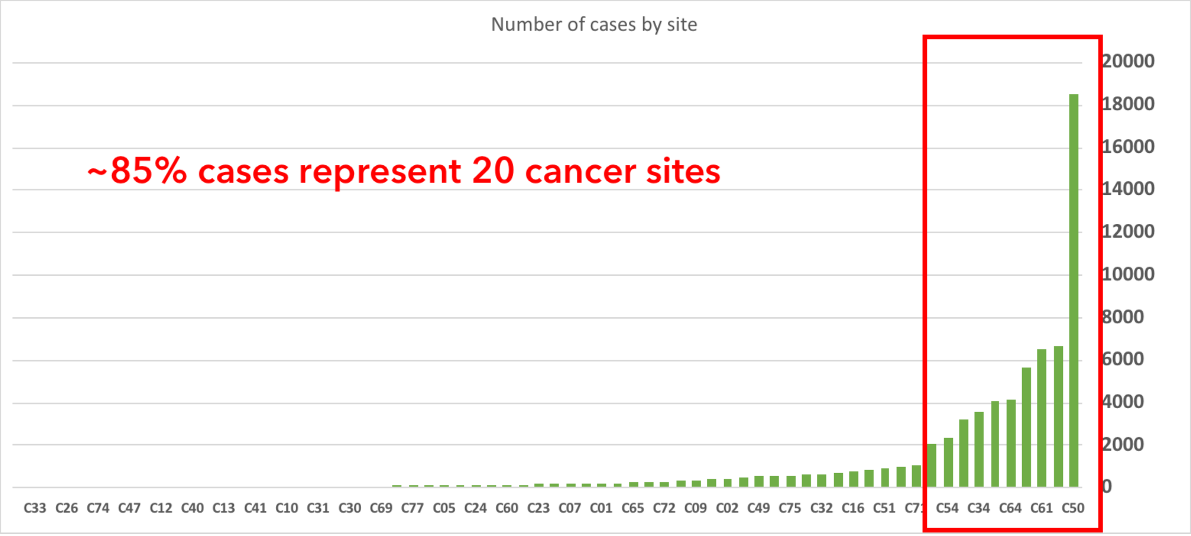


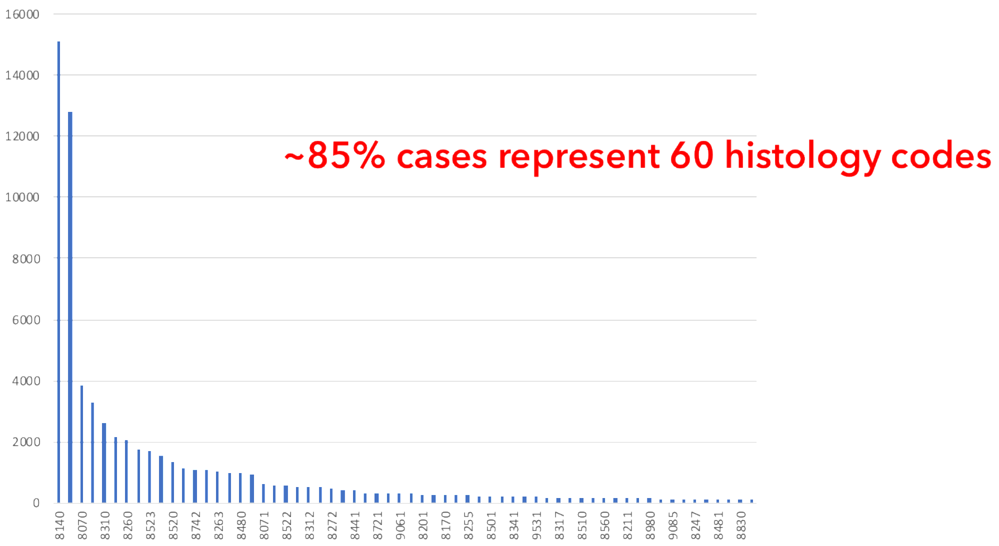
Louisiana Tumor Registry

- 2004-2018
- 374,826 pathology documents

Kentucky Cancer Registry

- 2004-2018
- 171,890 pathology documents





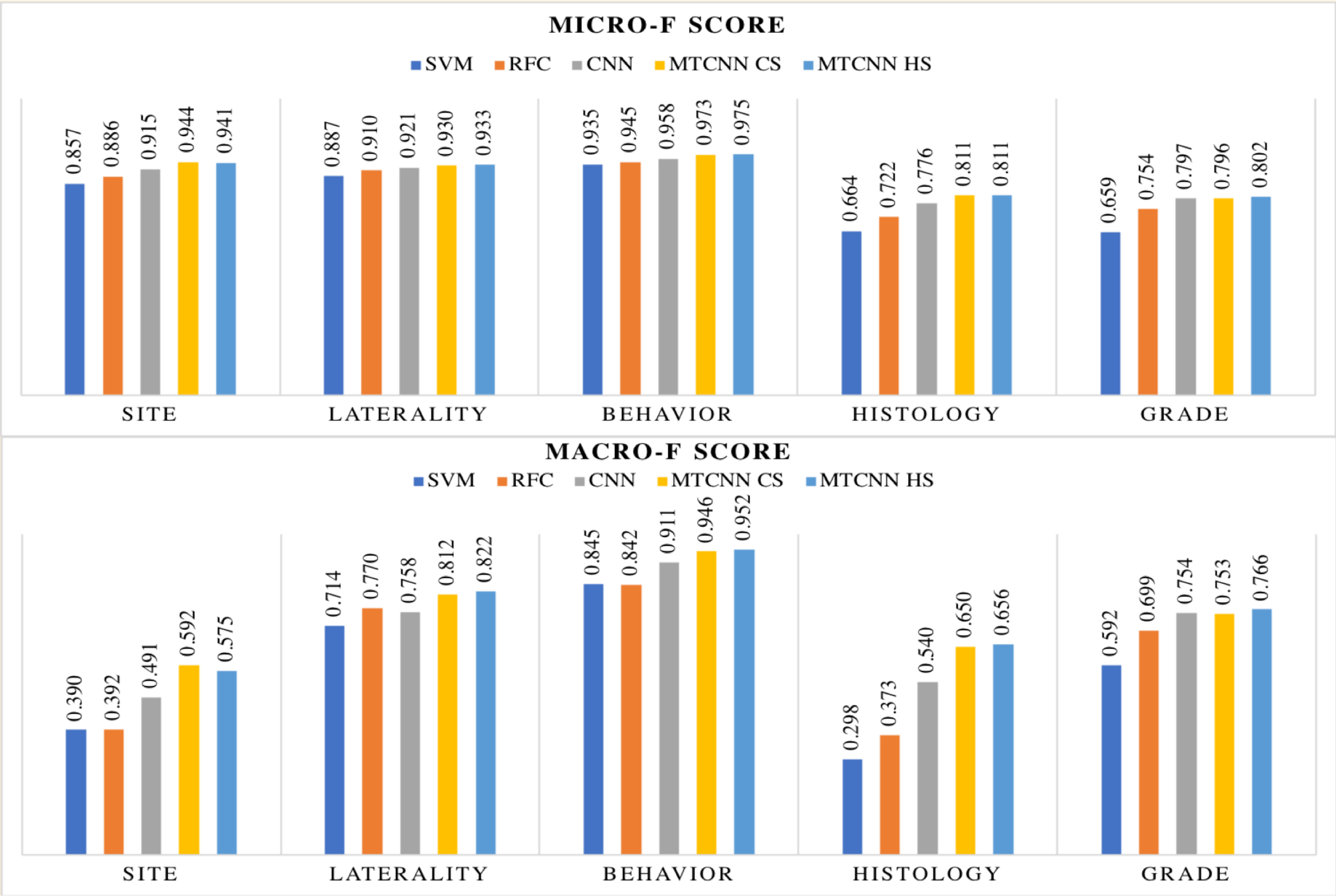
11

Performance Metrics

- Precision = TP/(TP+FP)
  - Precision=PPV
- Recall = TP / (TP+FN)
  - Recall=Sensitivity
- F1 = (2xPre x Rec) / (Pre + Rec)
  - A measure that combines both precision and recall

- Macro-Averaging
  - Average all Pre/Rec/F1 values
  - *i.e., all classes are weighted equally*
- Micro-Averaging
  - Sum up classification decisions for each case
  - Calculate Pre/Rec/F1 from the summations
  - *i.e., all cases are weighted equally*

2-fold Cross-Validation results



SVM: Support Vector Machine
RFC: Random Forest
CNN: Single-task Convolutional Neural Network
MT-CNN (CS): Multi-task CNN (cross stitch)
MT-CNN (HP): Multi-task CNN (hard parameter sharing)

API Deployment and Testing: across 11 SEER Registries / ~3M docs

micro-F1 scores

REGISTRY	A	B	C	D	E	F	G	H	I	J	K	Avg. across Registries	All Documents
Site	0.89	0.86	0.88	0.91	0.88	0.88	0.87	0.89	0.88	0.90	0.89	0.88	**0.885**
Histology	0.75	0.63	0.71	0.72	0.71	0.71	0.68	0.71	0.68	0.68	0.71	0.70	**0.702**
Laterality	0.89	0.88	0.90	0.88	0.89	0.88	0.88	0.88	0.88	0.85	0.88	0.88	**0.880**
Behavior	0.97	0.96	0.96	0.96	0.96	0.97	0.95	0.96	0.97	0.95	0.96	0.96	**0.960**
Grade	0.71	0.63	0.67	0.75	0.67	0.67	0.66	0.68	0.70	0.67	0.71	0.68	**0.684**
Total # of documents	381,316	40,535	322,462	236,933	166,247	79,634	923,086	345,118	172,153	234,493	125,568		**3,027,545**

42.5% correctly classified across (S+H+B+L+G)
31.2% correctly classified across (SS+H+B+L+G)

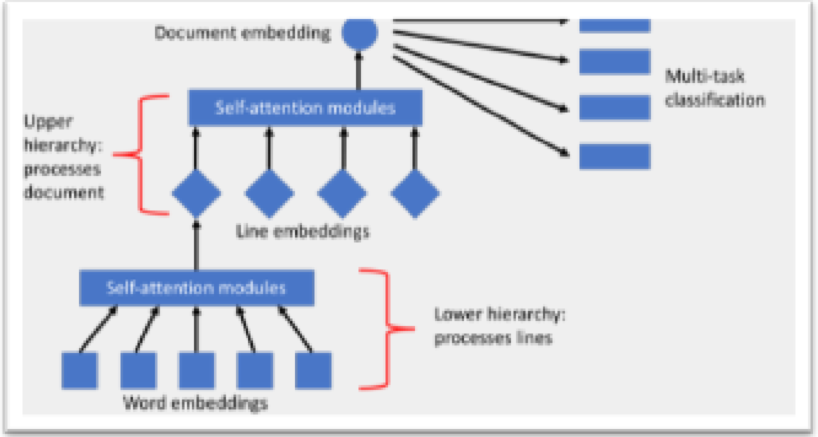
64.2% correctly classified across (S+H+B)
45.3% correctly classified across (SS+H+B)

U.S. DEPARTMENT OF ENERGY | NIH NATIONAL CANCER INSTITUTE

Impact on Cancer Registry Workflow

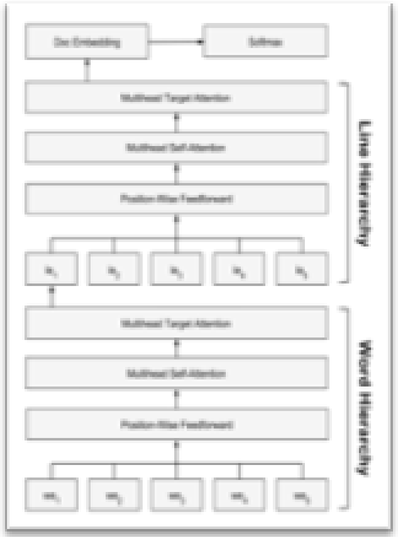
- Mean time for a registrar to code site, histology, behavior, grade, and laterality:
  - 55 seconds per clinical report
- Mean time for AI:
  - 12 milliseconds per report
- Real-world testing on 10 cancer registries and ~600K pathology reports (2018):
  - 4,048 hours for manual processing
  - 53 minutes with AI
- AI provides an opportunity for “real time” incidence reporting
  - goal to report at beginning of calendar year for prior calendar year (within 2-3 years)

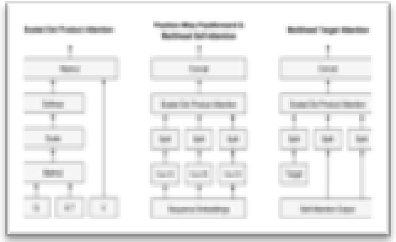
Newer NLP models, continuing to improve performance



Multi-Task Hierarchical Convolutional Attention Network (MT-HCAN)

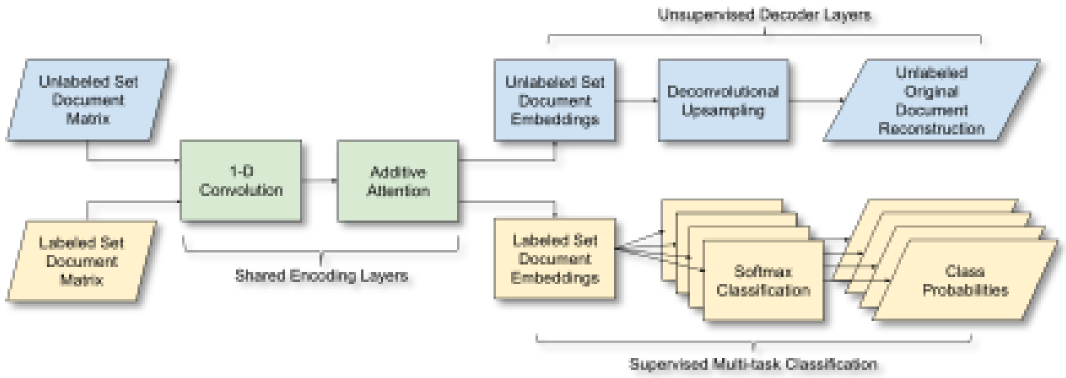
Gao, S. et al. "Hierarchical Convolutional Attention Networks for Text Classification." Proceedings of The 3rd Workshop on Representation Learning for NLP, pp. 11-23 2018. http://www.aclweb.org/anthology/W18-3002





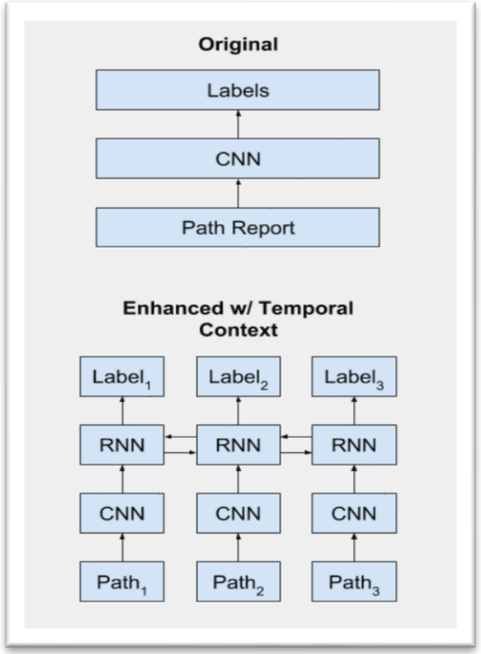
Hierarchical Self-Attention Network (HiSAN)

Gao, S. et al. "Classifying Cancer Pathology Reports with Hierarchical Self-Attention Networks." (submitted to Nature Medicine)



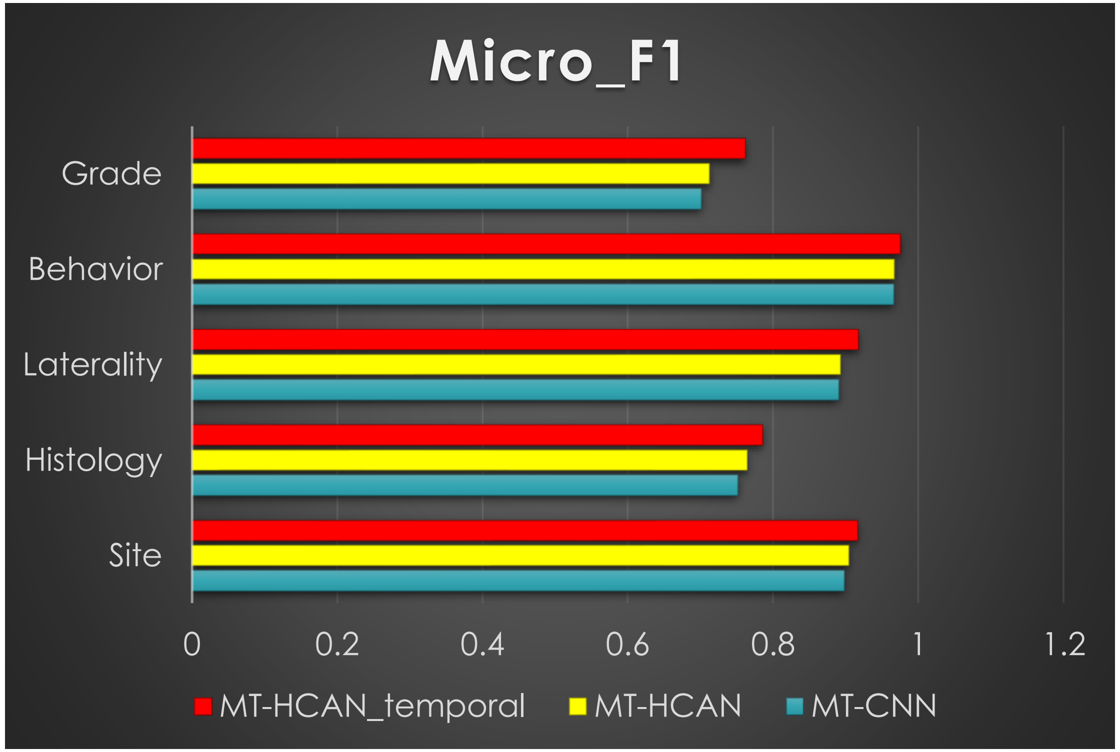
Semi-Supervised Multi-Task Attention CNN

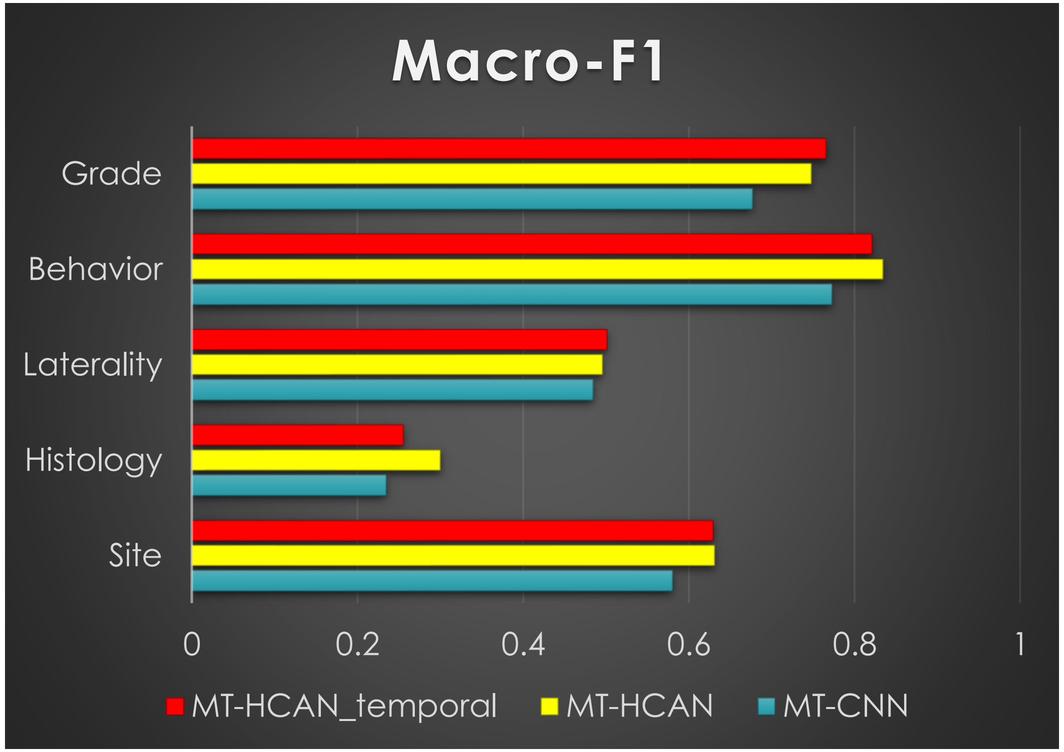
Qiu J.X. et al. "Semi-Supervised Information Extraction for Cancer Pathology Reports." 2019 IEEE Biomedical and Health Informatics Conference (submitted)



MT-HCAN-RNN for longitudinal analysis of text documents records

Preliminary Results





Software Deployment via DOE's CANDLE framework

- **Can**cer **D**istributed **L**earning **E**nvironment (CANDLE) Program
  - An exscale deep learning environment for cancer research
  - Building on open source Deep Learning frameworks
  - Collaboration between DOE computing centers, HPC vendors and ECP co-design and software technology projects
- ECP-CANDLE GitHub: https://github.com/ECP-CANDLE
- ECP-CANDLE FTP Site: http://ftp.mcs.anl.gov/pub/candle/public/

Software release via JDACS4C IP Committee Repository

- Multi-Task Convolutional Neural Networks (https://github.com/CBIIT/jdacs4c-staging/tree/master/ORNL_MT-CNN)

- PathRepHan (https://github.com/CBIIT/jdacs4c-staging/tree/master/PathRepHAN)

Conclusions & Next Steps

- Deep learning shows promise for automated information extraction from unstructured pathology reports to increase efficiency, data quality, and timeliness of cancer surveillance.
- MT-CNN performance exceeded that of traditional ML and single-task CNNs
- Our hard-parameter sharing MT-CNN is capable of scaling effectively across documents and information extraction tasks without additional computational or domain expert demands.
- Cross-registry performance remained fairly robust across all tasks.
- Other DL methods in the pipeline
- Human-AI integration is an open-ended question
    - What is the most effective way to integrate AI in national cancer surveillance?
    - Is interpretability possible and/or important?
    - Case-level uncertainty quantification maybe helpful

Final Thoughts on AI for Health

- Hope
  - The convergence of big data and AI will enable the accumulation and automation of functional knowledge in biomedicine
- Hype
  - AI solutions are superior to collective intelligence of the experts
  - Practical translation of AI tools is straightforward
- Hard Truth
  - Need for sustainable infrastructure to democratize AI innovation
  - Need for scalable algorithms to support the continuum of scientific discovery and clinical application
  - Human-AI integration approach will impact real-world value
  - AI interpretability and (real-time) uncertainty quantification are important future directions
  - Vulnerability issues for AI models and AI users (cognitive hacking) are critical

ACKOWLEDGEMENTS

This work has been supported in part by the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) program established by the U.S. Department of Energy (DOE) and the National Cancer Institute (NCI) of the National Institutes of Health. This work was performed under the auspices of the U.S. Department of Energy by Argonne National Laboratory under Contract DE-AC02-06-CH11357, Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, Los Alamos National Laboratory under Contract DE-AC5206NA25396, and Oak Ridge National Laboratory under Contract DE-AC05-00OR22725.

The authors gratefully acknowledge the contributions of the state and regional cancer registry staffs for their work in collecting the data used in this study.

THANK YOU!!!