Oak Ridge National Laboratory

DOE-NCI AI Activities in Cancer Surveillance Applicable to NCCR

Georgia Tourassi, PhD

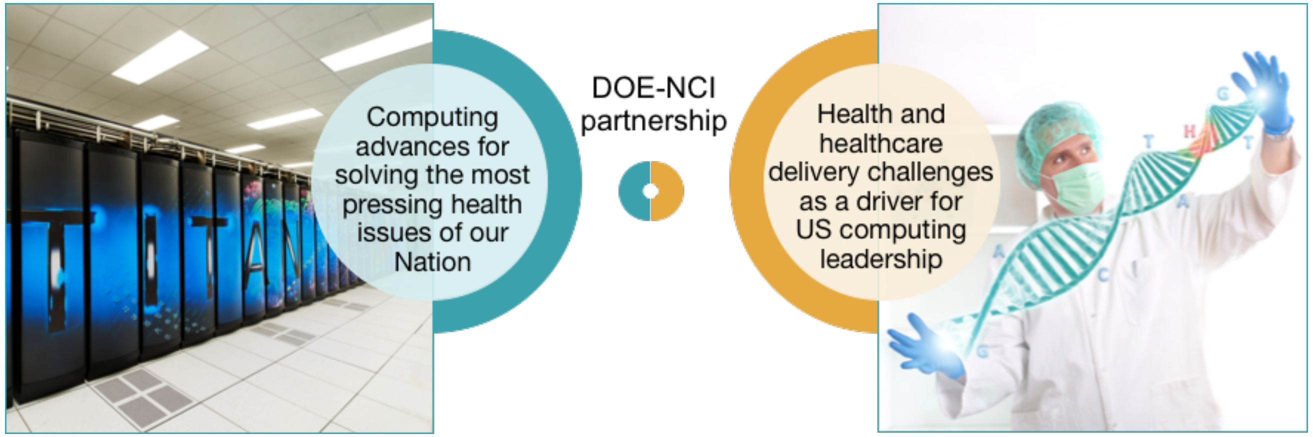
Director, National Center for Computational Sciences

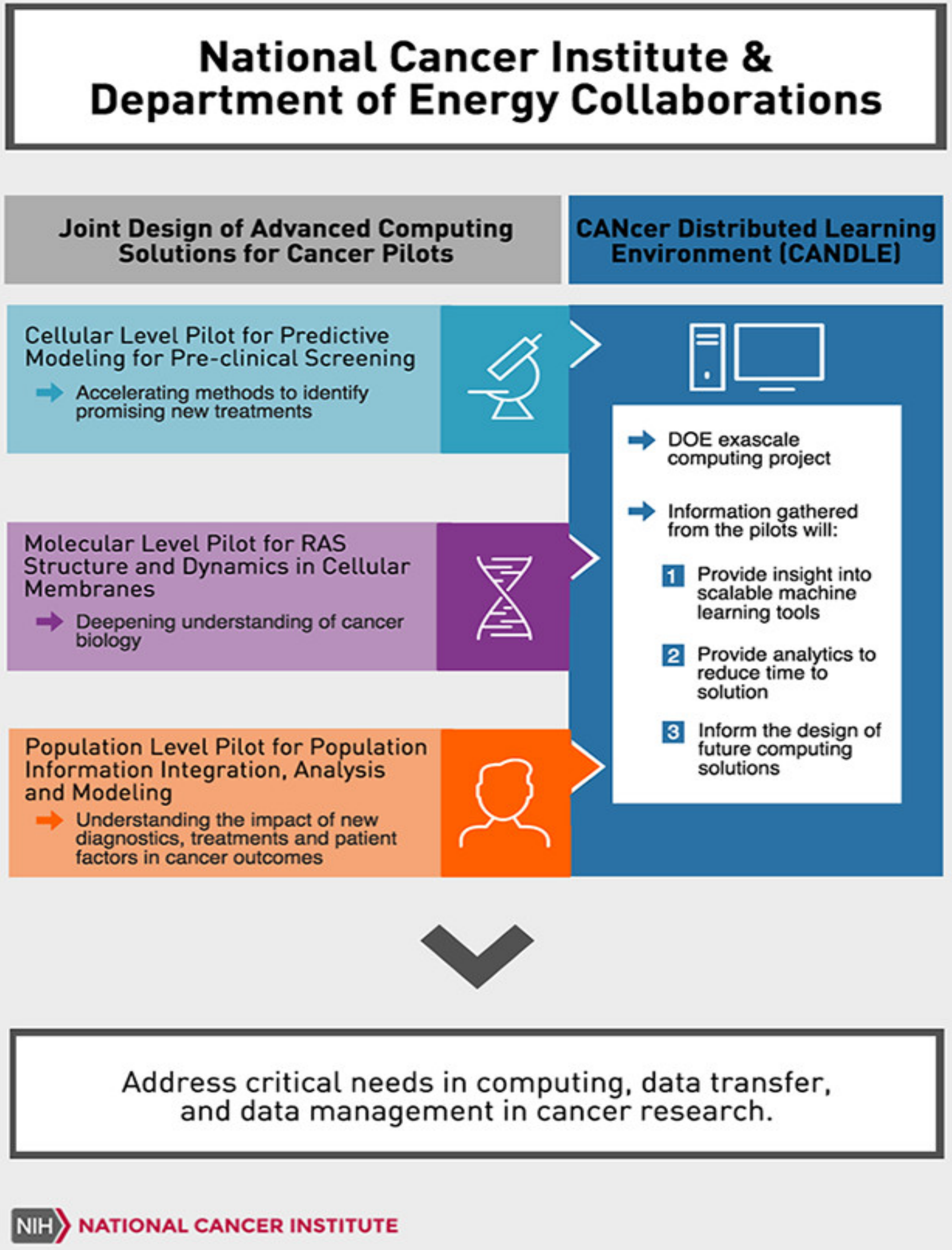
Oak Ridge National Laboratory

Presented at the NCCR Data Summit

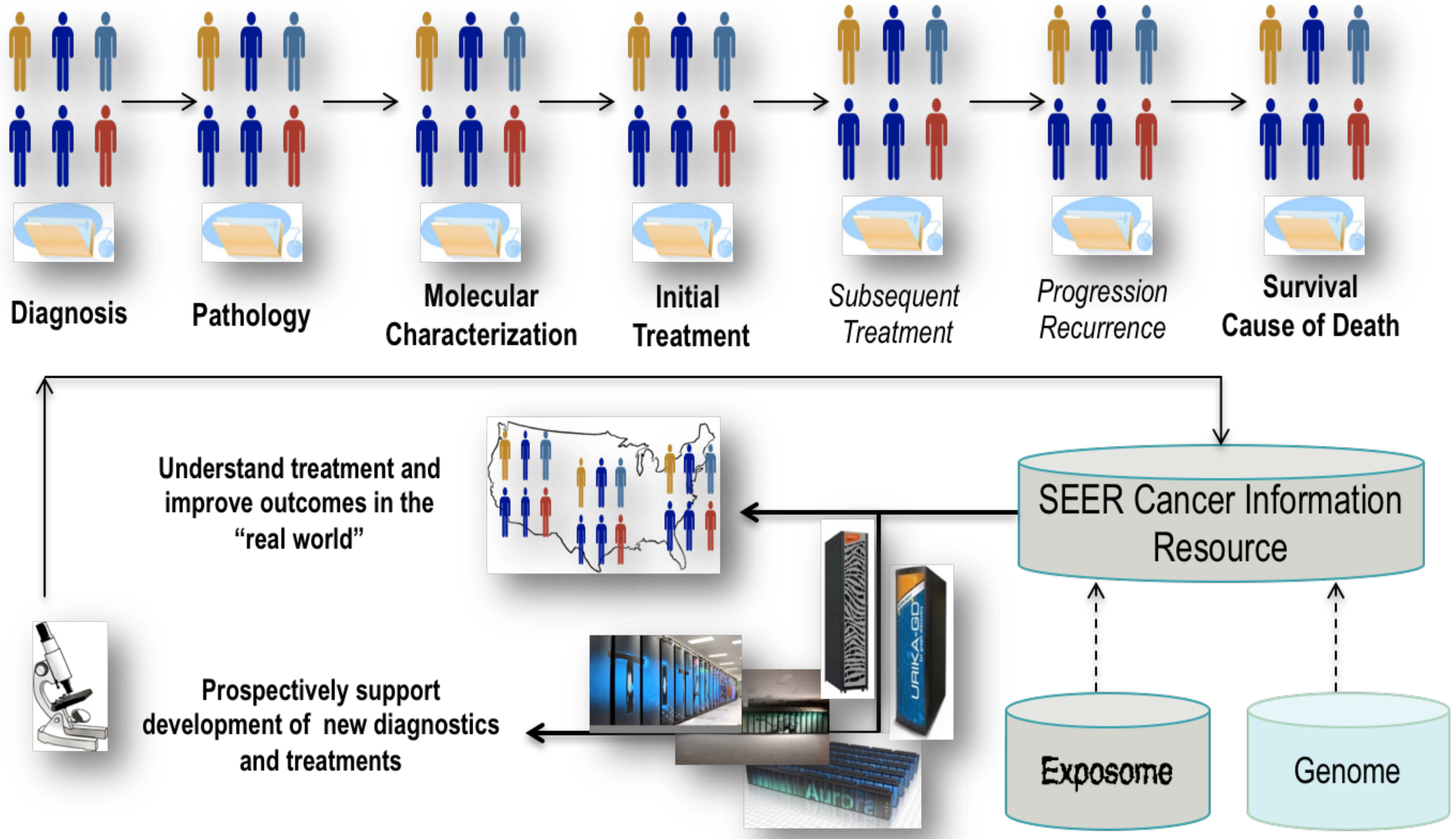
February 8, 2021

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



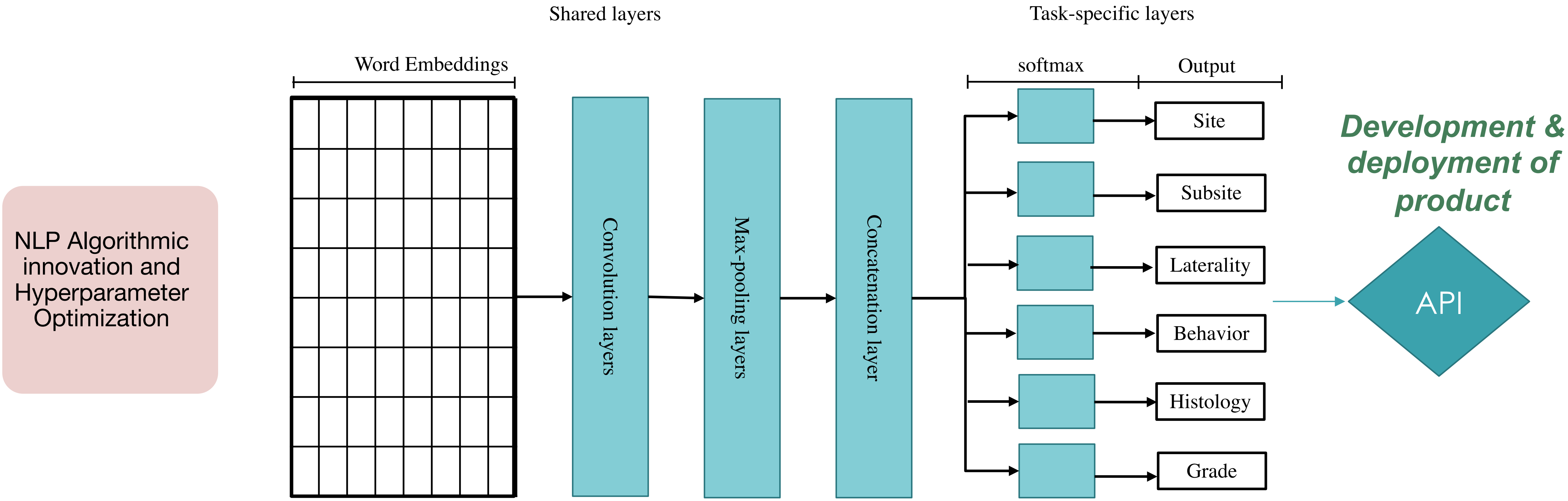


AI to Modernize the National Cancer Surveillance Program



To develop and deploy robust and scalable AI solutions for automated information capture from free text pathology reports.

Information Extraction of Reportable Data Elements



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

>20,000 cancer phenotypes observed in based only on 6 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 → Exascale computing**

Iterative Improvement and Testing: 13 SEER Registries, ~4M documents

Trained on 2 registries

V6	AA	BB	CC	DD	EE	FF	GG	HH	II	JJ	KK	LL	MM	Average
Site	93.96%	92.04%	93.10%	94.54%	92.84%	92.64%	95.23%	93.13%	93.97%	93.86%	94.38%	93.95%	93.41%	93.62%
Histology	84.52%	79.33%	82.67%	83.03%	84.11%	82.50%	85.63%	82.86%	82.68%	81.03%	80.19%	82.22%	78.82%	82.28%
Laterality	93.38%	92.39%	93.92%	92.95%	93.28%	93.68%	94.76%	93.06%	92.93%	94.24%	92.31%	94.22%	90.39%	93.19%
Behavior	96.61%	96.47%	95.81%	96.51%	95.97%	96.84%	97.52%	95.25%	96.40%	97.23%	95.58%	96.53%	97.45%	96.47%
Grade	79.82%	75.23%	77.06%	81.20%	78.83%	78.66%	82.93%	79.38%	78.16%	78.15%	79.92%	81.55%	79.12%	79.23%
Average	89.66%	87.09%	88.51%	89.65%	89.01%	88.86%	91.21%	88.74%	88.83%	88.90%	88.48%	89.69%	87.84%	

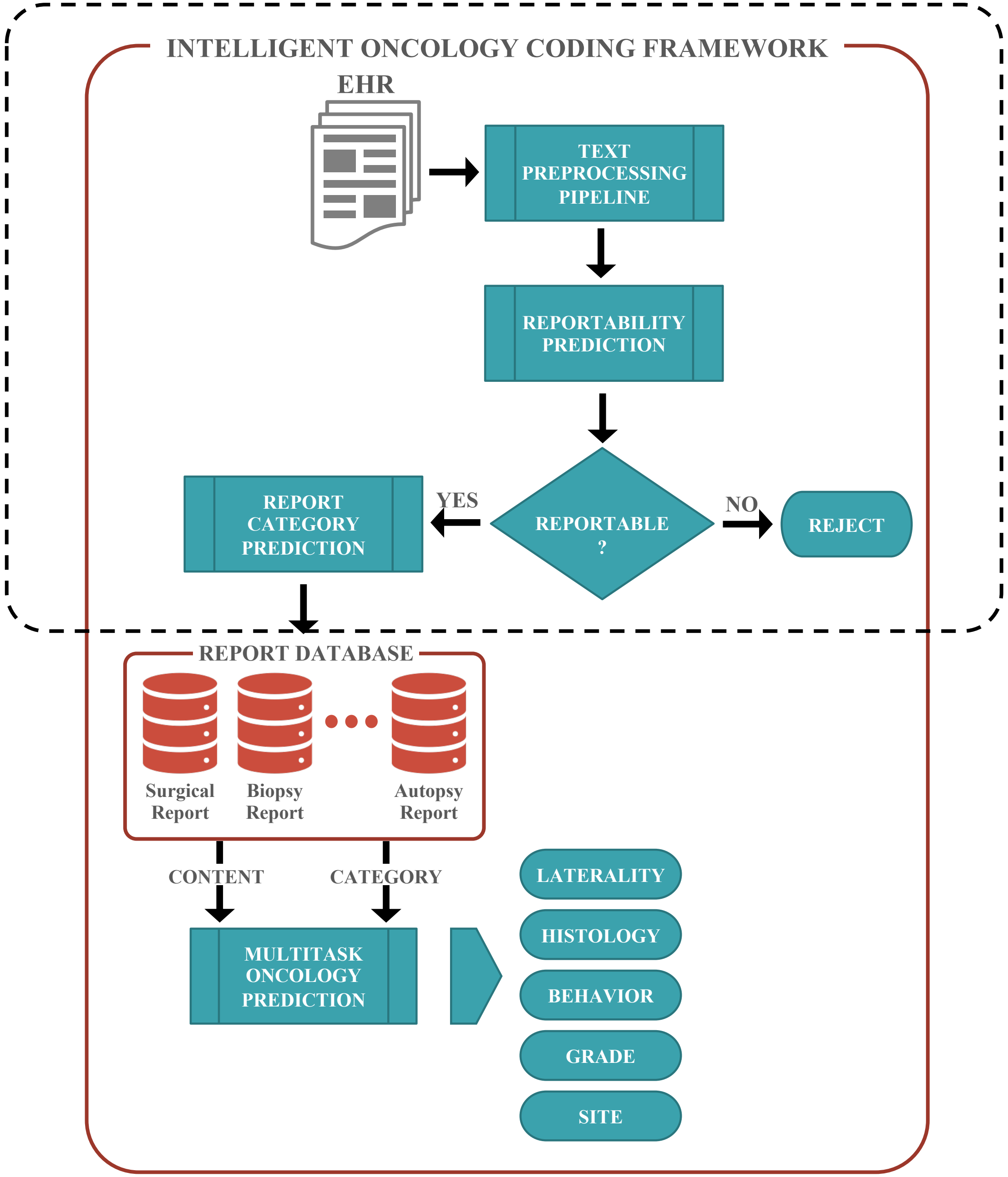
Trained on 4 registries

V7	AA	BB	CC	DD	EE	FF	GG	HH	II	JJ	KK	LL	MM	Average
Site	93.88%	91.89%	92.95%	94.48%	93.18%	92.69%	95.37%	93.15%	94.55%	93.91%	94.11%	95.11%	92.67%	93.69%
Histology	87.23%	83.58%	87.12%	86.26%	87.90%	86.33%	88.80%	87.09%	86.91%	84.74%	83.04%	87.56%	83.91%	86.19%
Laterality	94.21%	92.96%	94.11%	93.92%	93.95%	93.99%	95.27%	93.94%	94.02%	94.37%	93.55%	95.27%	92.32%	93.99%
Behavior	96.67%	96.71%	96.22%	96.94%	96.69%	97.10%	97.74%	95.74%	96.96%	97.58%	95.76%	97.17%	97.35%	96.82%
Grade	81.60%	80.21%	78.99%	83.52%	81.13%	81.95%	85.15%	80.47%	81.87%	80.44%	82.77%	85.23%	82.67%	82.00%
Average	90.72%	89.07%	89.88%	91.02%	90.57%	90.41%	92.47%	90.08%	90.86%	90.21%	89.85%	92.07%	89.78%	

- **Manual** path screening for 5 variables – 1 year,
  - 600,000 path reports: 4,048 hrs (**55sec/report**)
- **AI** path screening on same task: 55 min (**12msec/report**)
- **4500x speed gain to enable near real time cancer surveillance**

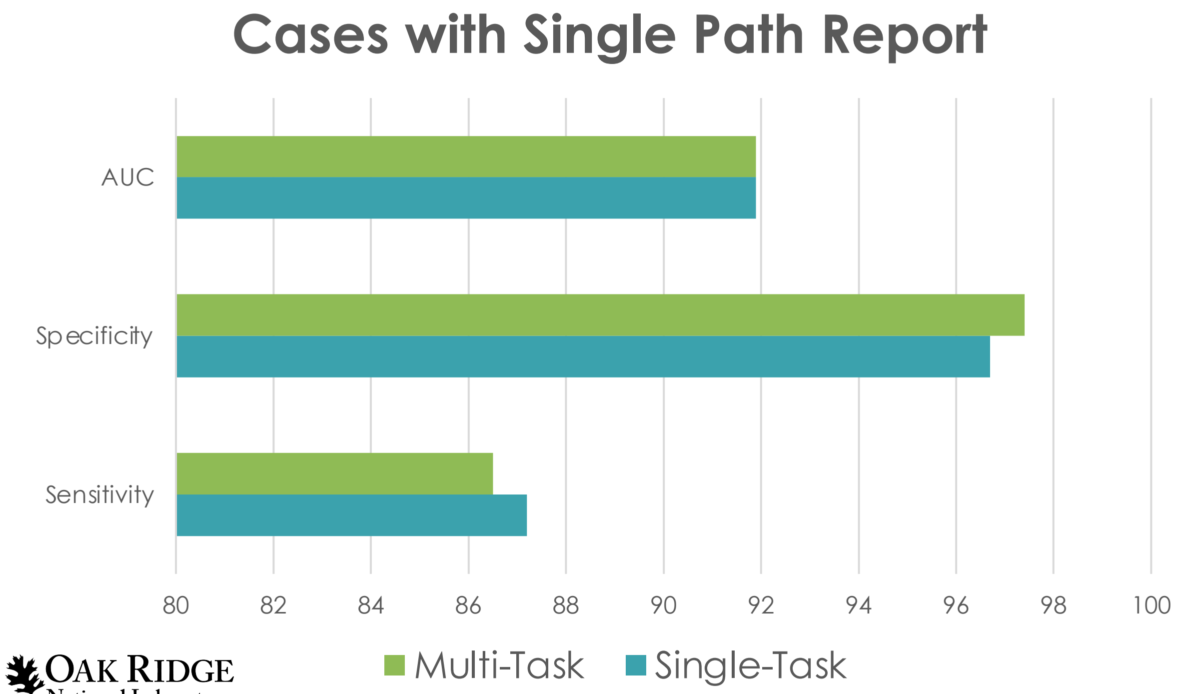
Reportability

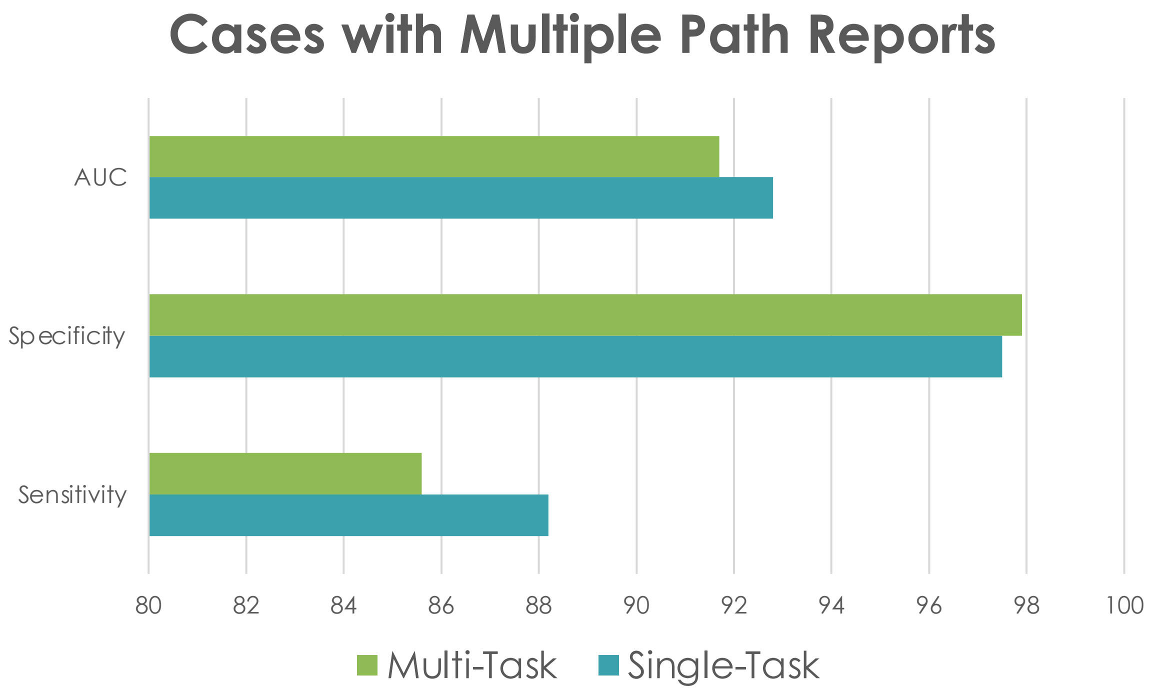
- "Reportability" assessment is currently being done either manually or by software at path labs
- Even for pre-screened "reportable" path reports- 25-60% are actually "non-reportable"
- Two approaches explored:
  - Predictive model to automatically determine the reportability status of unstructured pathology reports for labs without pre-screening software
  - Predictive model to differentiate pre-screened reportable path reports at the registry (reducing the need to manually review the 25-60% FPs)
- Deep learning models trained with data from different registries achieve cross-validated classification performance of 92%-99% depending on registry and cross-validation scenario.



Recurrence

- Identification of recurrent metastatic disease is a critical outcome for both patients and providers
- Recurrence also a key focus of many clinical trials
- <u>Initial Objective:</u> Identify pathology reports indicating “de novo” met
- <u>Hypothesis:</u> Model trained to detect metastasis at the time of diagnosis (for which we have CTC gold standard) can be used to detect metastasis indicative of disease progression.





Translation of deep learning NLP models developed to support SEER operations on childhood cohort

- First step: Information extraction DL models trained with both adult and pediatric cancer cases appear to be consistently better than the pediatric-cohort model
  - Incorporating adult cases boosts performance
  - <u>Transfer learning from adult cases to childhood cases is helpful (boosting training set)</u>
- Since prevalence of pediatric cancers differs, we may lose details if we simply apply “all cases model” as-is
  - Leukemia, Lymphoma and CNS Tumors are the most common
  - Adjusting the ratio of adult:pediatric training cases helps boost performance (considering the pediatric cancers are rare, ~2% of all cancers)
  - <u>Better to collect more pediatric cancer cases to train specialized models</u>
- The International Classification of Childhood Cancer (ICCC) is based on tumor morphology and primary site with an emphasis on morphology rather than the emphasis on primary site for adults.

International Classification of Childhood Cancer (ICCC)

	Accuracy	Model 1 (MT-CNN, inferred ICCC)	Model 2 (ST-CNN trained with ICCC ground truth)
Main Group (13 ICCC classes)	Micro F1	0.9467	0.9519
	Macro F1	0.8987	0.9248
Subgroup (47 ICCC classes)	Micro F1	0.8960	0.9078
	Macro F1	0.7071	0.7661

- Decent results from the SEER API recoding to ICCC
- <u>BUT – notable improvement with a specialized pediatric model trained with ICCC code labels</u>

Oak Ridge National Laboratory

Next Steps

- Leveraging
  - reportability API from path reports to radiology reports focusing on underreported childhood cancers (CNS/brain)
  - recurrence API based on path reports to radiology reports to identify metastatic recurrence from radiology reports for childhood cancers
- Development and training of DL NLP algorithm to extract structured treatment information from unstructured text documentation for childhood cancers
- Begin development of NLP models from radiology/path reports for extracting data to perform longitudinal capture of disease progression for pediatric cancers using PRISSMM annotated data

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!!!

tourassig@ornl.gov