Evaluating the Use of Artificial Intelligence in Oncology Diagnostics

Kenneth Hu, Pharm.D., Michele Bender, Pharm.D. Candidate, Sriravshini Kanukollu, Pharm.D. Candidate, Savan Patel, Pharm.D. Candidate, Evelyn Hermes-DeSantis, Pharm.D., BCPS, Michael Toscani, Pharm.D., Joseph A. Barone, Pharm.D., FCCP

Rutgers Institute for Pharmaceutical Industry Fellowships, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, NJ

Background

**Figure 1. Overview of Artificial Intelligence (AI)**

- Machine Learning: Supervised Learning, Unsupervised Learning, Reinforcement Learning
- Natural Language Understanding: Information Extraction, Ontology Engineering, Speech Processing, Statistical Semantics
- Deep Learning: Computer Vision, Speech Recognition, Natural Language Processing, Machine Translation

**ARTIFICIAL INTELLIGENCE**

Objective

- To evaluate four clinical studies which used artificial intelligence in oncology diagnostics.

Methods

- The Oncology Business Review was surveyed for recent studies of oncology diagnostics.
- Four studies of four cancer types published from August 2018-March 2019 were identified.

Results

**Figure 2. Overview of Studies**

**LUNG STUDY (Coudray N et al.)**

- **CANCER TYPE:** Lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) are the two most common subtypes, with treatment options differing for the two.
- **STRENGTHS:** Models maintained performance when tested on independent datasets of frozen and formalin-fixed paraffin-embedded tissues and on images from biopsies.
- **HIGHLIGHTS:** Trained deep convolutional neural network to accurately and automatically classify LUAD, LUSC, or normal lung tissue.
- **FUTURE OPPORTUNITIES:** Images used to train the deep neural network may not fully represent the diversity and heterogeneity of tissues pathologists typically inspect.
- **Performance was comparable to pathologists',** including predicting six of the most commonly mutated genes in LUAD from pathology images.

**OVARIAN STUDY (Lu H et al.)**

- **CANCER TYPE:** 7th most common form of cancer in women, and 8th leading cause of deaths from cancer.
- **HIGHLIGHTS:** Machine learning tested for reliability and reproducibility.
- **FUTURE OPPORTUNITIES:** Two independent validation datasets used to confirm overall survival differences among 8 patient groups stratified by RPV.

**BREAST STUDY (Lehman CD et al.)**

- **CANCER TYPE:** Most commonly diagnosed cancer in United States.
- **HIGHLIGHTS:** Deep Learning (DL) model developed and tested with 58,894 randomly selected digital mammograms.
- **FUTURE OPPORTUNITIES:** No mammographic examinations excluded, allowing for more inclusive pool to more accurately measure breast density in a multitude of women.

**STRENGTHS:**

- Models maintained performance when tested on independent datasets of frozen and formalin-fixed paraffin-embedded tissues and on images from biopsies.
- Pathologists’ diagnoses were used as gold standard to assess performance.
- Trained deep convolutional neural network to accurately and automatically classify LUAD, LUSC, or normal lung tissue.
- Performance was comparable to pathologists’, including predicting six of the most commonly mutated genes in LUAD from pathology images.

Conclusions

- **STRENGTHS:**
  - Models maintained performance when tested on independent datasets of frozen and formalin-fixed paraffin-embedded tissues and on images from biopsies.
  - Pathologists’ diagnoses were used as gold standard to assess performance.
  - Trained deep convolutional neural network to accurately and automatically classify LUAD, LUSC, or normal lung tissue.
  - Performance was comparable to pathologists’, including predicting six of the most commonly mutated genes in LUAD from pathology images.

- **FUTURE OPPORTUNITIES:**
  - Images used to train the deep neural network may not fully represent the diversity and heterogeneity of tissues pathologists typically inspect.
  - Small number of slides containing positive gene mutations limited accuracy and could be improved with larger datasets.

- **STRENGTHS:**
  - Models maintained performance when tested on independent datasets of frozen and formalin-fixed paraffin-embedded tissues and on images from biopsies.
  - Pathologists’ diagnoses were used as gold standard to assess performance.
  - Trained deep convolutional neural network to accurately and automatically classify LUAD, LUSC, or normal lung tissue.
  - Performance was comparable to pathologists’, including predicting six of the most commonly mutated genes in LUAD from pathology images.

- **FUTURE OPPORTUNITIES:**
  - Images used to train the deep neural network may not fully represent the diversity and heterogeneity of tissues pathologists typically inspect.
  - Small number of slides containing positive gene mutations limited accuracy and could be improved with larger datasets.

- **STRENGTHS:**
  - Models maintained performance when tested on independent datasets of frozen and formalin-fixed paraffin-embedded tissues and on images from biopsies.
  - Pathologists’ diagnoses were used as gold standard to assess performance.
  - Trained deep convolutional neural network to accurately and automatically classify LUAD, LUSC, or normal lung tissue.
  - Performance was comparable to pathologists’, including predicting six of the most commonly mutated genes in LUAD from pathology images.

- **FUTURE OPPORTUNITIES:**
  - Images used to train the deep neural network may not fully represent the diversity and heterogeneity of tissues pathologists typically inspect.
  - Small number of slides containing positive gene mutations limited accuracy and could be improved with larger datasets.

- **STRENGTHS:**
  - Models maintained performance when tested on independent datasets of frozen and formalin-fixed paraffin-embedded tissues and on images from biopsies.
  - Pathologists’ diagnoses were used as gold standard to assess performance.
  - Trained deep convolutional neural network to accurately and automatically classify LUAD, LUSC, or normal lung tissue.
  - Performance was comparable to pathologists’, including predicting six of the most commonly mutated genes in LUAD from pathology images.

- **FUTURE OPPORTUNITIES:**
  - Images used to train the deep neural network may not fully represent the diversity and heterogeneity of tissues pathologists typically inspect.
  - Small number of slides containing positive gene mutations limited accuracy and could be improved with larger datasets.
Evaluating the Use of Artificial Intelligence in Oncology Diagnostics

Kenneth Hu, Pharm.D., Michele Bender, Pharm.D. Candidate, Sriravshini Kanukollu, Pharm.D. Candidate, Savan Patel, Pharm.D. Candidate, Evelyn Hermers-DeSantis, Pharm.D., BCPS, Michael Tociani, Pharm.D., Joseph A. Barone, Pharm.D., FCCP

Rutgers Institute for Pharmaceutical Industry Fellowships, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, NJ

Supplementary Results

Table 1. Overview of Artificial Intelligence1,3,8,9

<table>
<thead>
<tr>
<th>Study</th>
<th>Cancer Type</th>
<th>Highlights</th>
<th>Strengths</th>
<th>Computer Vision</th>
<th>Natural Language Processing</th>
<th>Deep Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung Study (Coudray N et al)</td>
<td>Lung adenocarcinoma (LUAD) and squamous cell carcinomas (LUSC) are two most common subtypes, with over 75,000 deaths and 100,000 new diagnoses per year in the United States.</td>
<td>• Trained deep convolutional neural network (inception v3) on whole-slide and fixed paraffin-embedded (FFPE) tissue images from The Cancer Genome Atlas (TCGA) that has strong prognostic power and is also noninvasive measured by radiomics.</td>
<td>• Validated on independent datasets of frozen tissues, formalin fixed paraffin-embedded (FFPE) tissues, and biopsies.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung Study (Coudray N et al)</td>
<td>Lung adenocarcinoma (LUAD) and squamous cell carcinomas (LUSC) are two most common subtypes, with over 75,000 deaths and 100,000 new diagnoses per year in the United States.</td>
<td>• Trained deep convolutional neural network (inception v3) on whole-slide and fixed paraffin-embedded (FFPE) tissue images from The Cancer Genome Atlas (TCGA) that has strong prognostic power and is also noninvasive measured by radiomics.</td>
<td>• Models maintained performance when tested on datasets of frozen and FFPE tissue and on images from biopsies.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung Study (Coudray N et al)</td>
<td>Lung adenocarcinoma (LUAD) and squamous cell carcinomas (LUSC) are two most common subtypes, with over 75,000 deaths and 100,000 new diagnoses per year in the United States.</td>
<td>• Trained deep convolutional neural network (inception v3) on whole-slide and fixed paraffin-embedded (FFPE) tissue images from The Cancer Genome Atlas (TCGA) that has strong prognostic power and is also noninvasive measured by radiomics.</td>
<td>• Pathologists’ diagnoses were used as gold standard to assess performance of study approach.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast Study (Lehman CD et al)</td>
<td>Most commonly diagnosed cancer in the United States. 1,2,4,5</td>
<td>• Retrospective study using 364 EOC epithelial ovarian cancer patients, validated in two independent, multi-center cohorts.</td>
<td>• Two independent validation datasets (The Cancer Genome Atlas (TCGA) validation and Hammersmith Hospital (HH) validation cohorts) were used to confirm overall survival (OS) differences amongst 3 patient groups stratified based on RPV.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast Study (Lehman CD et al)</td>
<td>Most commonly diagnosed cancer in the United States. 1,2,4,5</td>
<td>• Retrospective study using 364 EOC epithelial ovarian cancer patients, validated in two independent, multi-center cohorts.</td>
<td>• Correlation index (op score) used to measure that RPV improved clinically available prognostic methods in all 3 datasets.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast Study (Lehman CD et al)</td>
<td>Breast cancer is the most commonly occurring cancer in women. 1,2,5,7</td>
<td>• Retrospective study using 364 EOC epithelial ovarian cancer patients, validated in two independent, multi-center cohorts.</td>
<td>• Biological interpretation of RPV was evaluated using Spearman correlation coefficients of gene expression correlation used to confirm consistent feature-wise correlation in HH and TCGA cohorts, which is an indicator of high reliability (r = 0.817, p &lt; 0.0001).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate Study (Donovan MJ et al)</td>
<td>Most commonly diagnosed solid tumor in men. 1,2,4,5</td>
<td>• Developed and tested deep learning (DL) model by using 588,984 randomly selected digital mammograms.</td>
<td>• No mammographic examinations excluded (i.e. exclusions due to prior surgery, implants, etc.) allowing for more inclusive pool of references to more accurately measure breast density in all populations.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate Study (Donovan MJ et al)</td>
<td>Most commonly diagnosed solid tumor in men. 1,2,4,5</td>
<td>• Developed and tested deep learning (DL) model by using 588,984 randomly selected digital mammograms.</td>
<td>• No mammographic examinations excluded (i.e. exclusions due to prior surgery, implants, etc.) allowing for more inclusive pool of references to more accurately measure breast density in all populations.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Overview of Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Cancer Type</th>
<th>Highlights</th>
<th>Strengths</th>
<th>Future Opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung Study (Coudray N et al)</td>
<td>Lung adenocarcinoma (LUAD) and squamous cell carcinomas (LUSC) are two most common subtypes, with over 75,000 deaths and 100,000 new diagnoses per year in the United States.</td>
<td>• Trained deep convolutional neural network (inception v3) on whole-slide and fixed paraffin-embedded (FFPE) tissue images from The Cancer Genome Atlas (TCGA) that has strong prognostic power and is also noninvasive measured by radiomics.</td>
<td>• Future work on deep-learning model visualization tools would help identify and characterize features used by neural network.</td>
<td>• Retrospective design, future prospective study or analysis of retrospective randomized clinical trial data is required to validate RPV in a more general HGSOC population.</td>
</tr>
</tbody>
</table>
| Breast Study (Lehman CD et al) | Breast cancer is the most commonly occurring cancer in women. 1,2,5,7 | • Retrospective study using 364 EOC epithelial ovarian cancer patients, validated in two independent, multi-center cohorts. | • Full study assessing RPV with both estro 


Evaluating the Use of Artificial Intelligence in Oncology Diagnostics

Kenneth Hu, Pharm.D., Michele Bender, Pharm.D. Candidate, Srivarshini Kanukollu, Pharm.D. Candidate, Savan Patel, Pharm.D. Candidate, Evelyn Hermer-DeSantis, Pharm.D., BCPS, Michael Toscani, Pharm.D., Joseph A. Barone, Pharm.D., FCCP

Rutgers Institute for Pharmaceutical Industry Fellowships, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, NJ

References