Pneumonia is a major cause of morbidity and mortality worldwide. Patients are typically diagnosed from a combination of chest X-rays (CXRs), clinical symptoms, and blood tests.

If we can diagnose pneumonia from CXRs using ML, we can help patients receive better and faster treatment by:

- reducing radiologist workload,
- reducing diagnosis cost, and
- accelerating time to diagnosis.

We collaborate with Intermountain Healthcare to update and improve CheXED, their currently deployed pneumonia detection model.

### Problem Statement

- We leverage real-world multimodal data (CXRs and radiology reports) from collaborators at Intermountain Healthcare to develop an image+text-SSL multi-task model for pneumonia, multilobar pneumonia, and pleural effusion.
- We simulate deploying this model in the healthcare system’s emergency departments in 2020, and “retrain” it with a regular cadence through 2021.
- We perform a detailed error analysis and demographic subgroup performance analysis to prepare our model for deployment.

### Dataset

- **CheXpert**: 224,316 CXRs and accompanying radiology reports from Stanford Medical Center from 2002-2017.
- **Intermountain**: 16,708 CXRs and accompanying radiology reports from Intermountain Emergency Departments and Urgent Cares from 2009-2021.

- **Train/val/test**: 15,774/467/467.
- **Preprocessing**: Random image crops and normalization.

### Methods and Results

**ImageNet-Supervised**

<table>
<thead>
<tr>
<th>Task</th>
<th>Supervised</th>
<th>CheXpert (10%)-Supervised</th>
<th>CheXpert-SSL (GLoRIA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumonia</td>
<td>Supervised</td>
<td>CheXpert</td>
<td>CheXpert-SSL (GLoRIA)</td>
</tr>
<tr>
<td>Effusion</td>
<td>Supervised</td>
<td>CheXpert</td>
<td>CheXpert-SSL (GLoRIA)</td>
</tr>
<tr>
<td>Multilobar</td>
<td>Supervised</td>
<td>CheXpert</td>
<td>CheXpert-SSL (GLoRIA)</td>
</tr>
<tr>
<td>Pleural Effusion</td>
<td>Supervised</td>
<td>CheXpert</td>
<td>CheXpert-SSL (GLoRIA)</td>
</tr>
</tbody>
</table>

**Global Representation Learning**

- GLoRIA++ is more robust to real-world “hard examples” than supervised CheXED.
- Many false positives and negatives arise from understandable mistakes.

### Conclusion and Next Steps

- **CheXpert-GLORIA++ CheXED achieves top or top-2 performance on all tasks.**
- **Differences between GLORIA and GLORIA++ task performance can be explained by the supervised task set, which indicates specific use cases for each model.**
- **We can maintain performance in the face of data drift via regular retraining after deployment. We improve performance for the pneumonia task with this method.**
- **Qualitative results indicate that GLORIA++ is more robust to real-world “hard examples” than the currently deployed model, and that GLORIA++’s mistakes are often on CXRs that radiologists would also find more difficult.**
- **Next steps: debias model, retrain with more recently-available data (both labeled and report-only), perform more few-shot and dataset size ablation studies to give annotation recommendations to collaborators.**