Deep learning frees us from feature engineering, but creates a new problem: “architecture engineering”. We use NAS to automate neural network design, with applications to novel scientific datasets.

**Datasets**
- **MNIST**: classifying images of handwritten digits
- **Graphene Kirigami**: cutting simulated graphene to optimize stress/strain properties
- **Galaxy Zoo**: classifying galaxy morphology from telescope images
- **Chest X-Ray**: predicting 15 diseases from chest x-ray images

**Results**

<table>
<thead>
<tr>
<th>Model</th>
<th>MNIST</th>
<th>Graphene</th>
<th>Galaxy Zoo</th>
<th>Chest X-Ray</th>
</tr>
</thead>
<tbody>
<tr>
<td>DARTS (Continuous)</td>
<td>99.07</td>
<td>0.89</td>
<td>0.094</td>
<td>0.157</td>
</tr>
<tr>
<td>DARTS (Discrete)</td>
<td>99.27</td>
<td>0.92</td>
<td>0.114</td>
<td>0.163</td>
</tr>
<tr>
<td>Random Search</td>
<td>99.31</td>
<td>0.90</td>
<td>0.098</td>
<td>0.169</td>
</tr>
<tr>
<td>ResNet</td>
<td>99.40</td>
<td>0.92</td>
<td>0.095</td>
<td>0.163</td>
</tr>
</tbody>
</table>

**Discussion**
- Architecture weights initialized to ~0.125
- Architecture considered sparse if many weights near 0
- Observation: degree of architecture sparsity varies considerably across datasets

**Conclusions & Future Work**
- DARTS is a useful tool, but overkill on simple tasks
- ResNet and random search could be good enough
- DARTS introduces many additional hyperparameters
- DARTS discretization step is heuristic
- Future work: encourage sparsity in DARTS architectures (e.g. sparsemax vs. softmax) to prevent discretization failure