- Accuracy on the Line (Miller)
  - https://arxiv.org/abs/2107.04649
- Measuring Robustness to Natural Distribution Shifts in Image Classification
  - https://arxiv.org/abs/2007.00644
- Effective Robustness against Natural Distribution Shifts for Models with Different Training Data (Accuracy on the Line follow-up)
  - https://arxiv.org/abs/2302.01381
- Model Performance Scaling with Multiple Data Sources (Hashimoto)
  - https://proceedings.mlr.press/v139/hashimoto21a.html
- Representation Matters: Assessing the Importance of Subgroup Allocations in Training Data (Rolf)
  - https://arxiv.org/abs/2103.03399
- Using Pre-Training Can Improve Model Robustness and Uncertainty
  - https://arxiv.org/abs/1901.09960
- An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (ViT)
  - https://arxiv.org/abs/2010.11929
- Vision Transformers are Robust Learners
  - https://arxiv.org/abs/2105.07581
- Robust physical-world attacks on deep learning visual classification (Eykholt)
  - https://arxiv.org/abs/1707.08945
- Robustness May Be at Odds with Accuracy (Tsipras)
  - https://arxiv.org/abs/1805.12152
- Information-theoretic analysis of generalization capability of learning algorithms (Xu and Raginsky)
  - https://arxiv.org/abs/1705.07809
- A Simple Framework for Contrastive Learning of Visual Representations (SimCLR)
  - https://arxiv.org/abs/2002.05709
- Learning Transferable Visual Models From Natural Language Supervision (CLIP)
  - https://arxiv.org/abs/2103.00020
- Language Models are Few-Shot Learners (GPT-3)
  - https://arxiv.org/abs/2005.14165