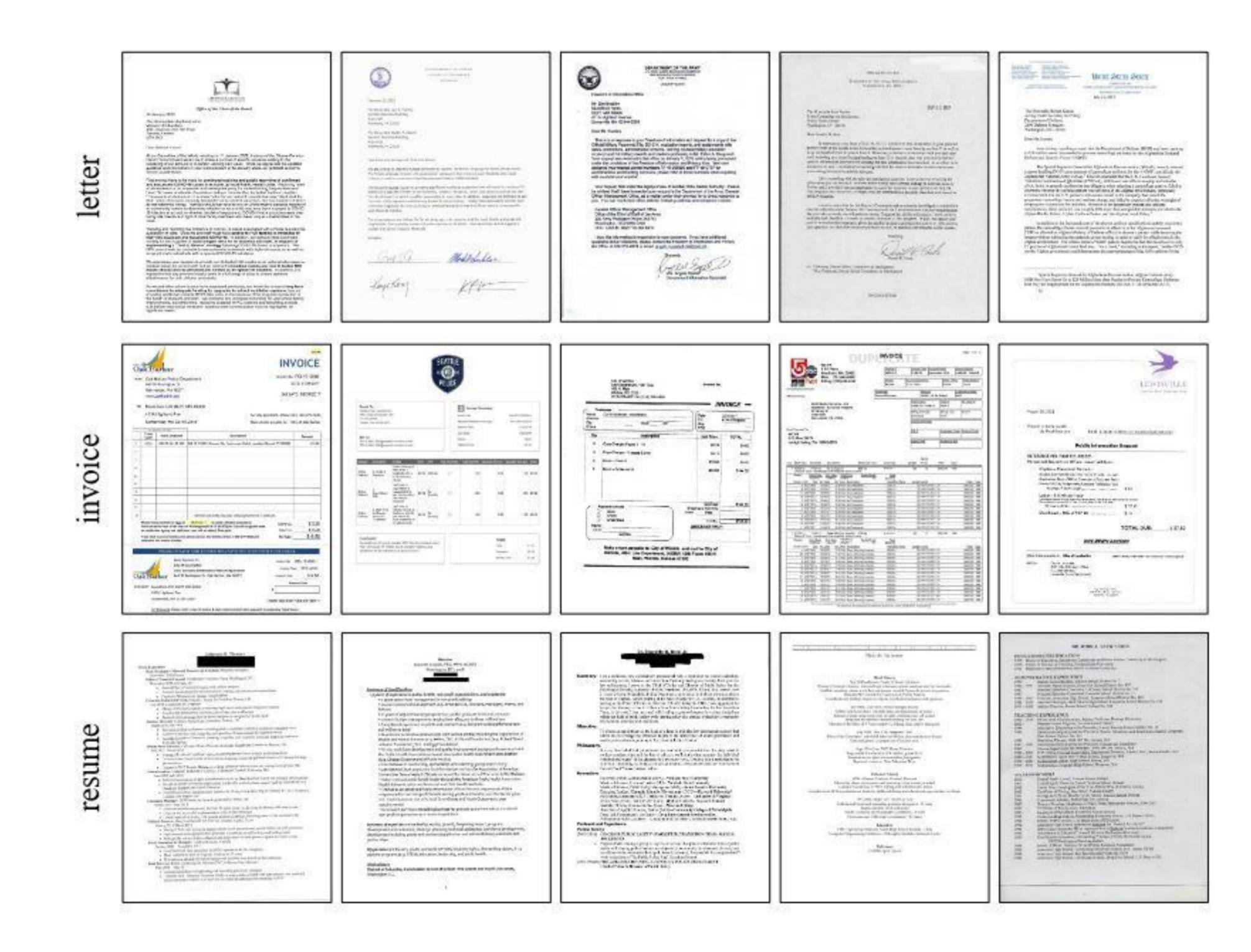# Out-of-Distribution Performance in Document Image Classification: Initial Findings

Stefan Larson[1], Gordon Lim[2], Yutong Ai[2], Brian Chen[2]
slarson@dryviq.com   1: DryvIQ, 2: University of Michigan

The RVL-CDIP corpus is the *de facto* standard benchmark for document classification, yet to our knowledge all studies that use this corpus do not include evaluation on *out-of-distribution (OOD)* documents. **In this work, we develop a new companion OOD dataset for evaluating document image classification performance on OOD documents.**

We create 2 new OOD benchmarks: OOD-*a*, consisting of documents that are *not* part of any of the 16 RVL-CDIP document categories, and OOD-*b*, new documents not from RVL-CDIP that *are* one of the 16 RVL-CDIP categories. Samples of OOD-*b* are shown at above.

Results of several image classifiers are shown below. **We observe that models tend to underperform on both OOD-*a* and OOD-*b* sets.** Our hope is that researchers incorporate OOD evaluation into their work using our new companion RVL-CDIP OOD benchmarks.

Model	ID Acc. (reported)	ID Acc. (achieved)	OOD-*a* AUC	OOD-*a* (Aug.)	OOD-*b* Acc.
VGG-16	0.910	0.905	0.885	0.870↓	0.683
ResNet-50	0.911	0.900	0.874	0.865↓	0.575
GoogLeNet	0.884	0.871	0.852	0.849↓	0.633
AlexNet	0.900	0.885	0.874	0.869↓	0.607
LayoutLMv2	0.953	0.887	0.843	0.832↓	0.533
CLIP (zs)	0.392	0.394	0.543	0.582↑	0.767