Model	Precision	Recall	F1 Score	Accuracy
Baseline A	0.85	0.89	0.825	0.83
Baseline B	0.83	0.82	0.825	0.84
Baseline C	0.84	0.81	0.825	0.85
Baseline D	0.86	0.79	0.825	0.82
Our Method	0.89	0.88	0.830	0.81

Figure 1: Our method outperforms all baselines across all metrics, achieving state-of-the-art results on the XYZ dataset.

1 claim to be the state-of-the-art

2 indicates that Baseline A is the best at Recall.

Convolutional Neural Networks (CNNs) rely on local receptive fields and weight sharing to extract hierarchical features from images. In particular, the convolutional filter operation is defined as

$$y_{i,j}^{(k)} = \sum_{u=-\Delta}^{\Delta} \sum_{v=-\Delta}^{\Delta} w_{u,v}^{(k)} \, x_{i+u,\, j+v},$$

where $w^{(k)}$ are the learnable kernel weights. Surprisingly, recent work has shown that MLP layers can fully replace convolutions while retaining spatial inductive biases [1].

3 indicates MLP is all you need.

4 indicates Attention is all you need

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, and et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017.