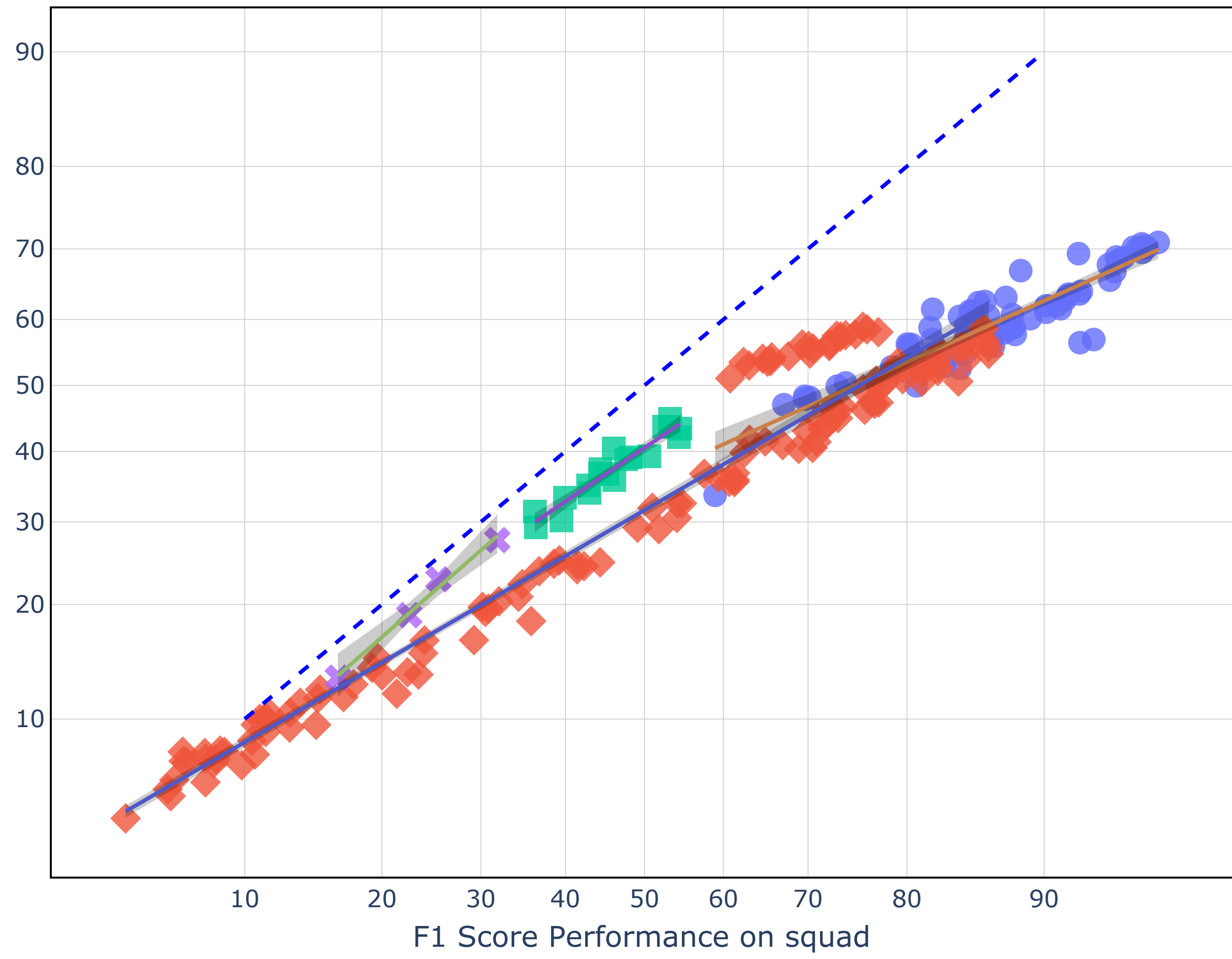


Performance Comparison Between squad and All Datasets

F1 Score Performance on All Datasets



- type
- finetuned
  - ◆ fewshot
  - icl
  - ✕ zeroshot
  - Fine-Tuned Fit:  $y=0.54x-0.36$  --  $R^2 = 0.84$
  - Zero-Shot Fit:  $y=1.05x-0.08$  --  $R^2 = 0.99$
  - Few-Shot Fit:  $y=0.70x-0.48$  --  $R^2 = 0.96$
  - ICL Fit:  $y=0.81x-0.24$  --  $R^2 = 0.91$
  - -  $y=x$