**Table S3.** Metrics obtained in the Kvasir-SEG test set for the different models and datasets used during training. Best value per metric is indicated in bold.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset for training/validation set** | **Network** | **Kvasir-SEG** | | | | | | |
| Accuracy | Precision | Recall | Specificity | F2-score | Jaccard | Dice |
| CVC-EndoSceneStill | U-Net+VGG16 | 87.63±13.25 | 67.59±32.87 | 61.51±33.74 | 94.68±8.62 | 57.64±31.54 | 46.29±29.36 | 57.24±30.72 |
| U-Net+Densenet121 | 90.53±14.05 | 77.80±32.68 | 63.31±37.22 | 98.15±6.23 | 62.72±36.38 | 56.12±34.29 | 64.26±35.35 |
| Linknet+VGG16 | 87.42±12.73 | 65.39±30.27 | 73.37±30.03 | 92.44±9.58 | 66.06±28.55 | 51.13±27.67 | 62.61±28.17 |
| Linknet+Densenet121 | 90.26±13.90 | 74.26±34.70 | 64.89±36.51 | **97.99±4.29** | 64.02±35.45 | 56.69±33.68 | 65.05±34.66 |
| Kvasir-SEG | U-Net+VGG16 | 93.34±8.66 | 79.93±27.73 | 75.04±32.90 | 97.77±3.85 | 73.30±31.08 | 64.23±29.78 | 72.94±29.76 |
| U-Net+Densenet121 | 94.64±7.75 | 81.03±21.93 | 88.89±22.25 | 96.29±5.89 | 85.36±21.58 | 74.13±23.40 | 82.28±21.76 |
| Linknet+VGG16 | 94.27±8.12 | **82.81±21.49** | 85.26±23.07 | 97.10±4.98 | 82.62±22.36 | 72.53±23.92 | 81.09±22.08 |
| Linknet+Densenet121 | **94.70±7.82** | 81.68±21.52 | **88.98±21.26** | 96.49±5.86 | **85.46±20.93** | **74.52±22.81** | **82.68±21.28** |
| PICCOLO | U-Net+VGG16 | 88.17±13.14 | 64.41±31.66 | 74.74±32.23 | 91.89±11.77 | 66.91±30.63 | 52.64±30.41 | 62.97±30.74 |
| U-Net+Densenet121 | 92.57±11.23 | 81.31±26.76 | 74.73±31.82 | 97.41±4.89 | 73.43±31.04 | 65.33±30.66 | 73.54±30.15 |
| Linknet+VGG16 | 91.02±10.94 | 72.60±30.55 | 74.73±32.22 | 95.49±7.08 | 70.55±30.58 | 58.99±30.40 | 68.54±30.28 |
| Linknet+Densenet121 | 91.29±13.63 | 75.91±31.17 | 70.37±35.96 | 95.74±11.03 | 68.09±34.81 | 59.61±33.80 | 67.58±34.14 |