|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Primary Origin | Precision | Recall | F1-score | mAP | AUC-ROC |
| Lung | **0.8571** | 0.94737 | 0.9 | 0.97051 | 0.99799 |
| Skin | 0.97561 | 0.90909 | 0.94118 | 0.99103 | 0.99836 |
| Kidney | 1 | **0.8571** | 0.92308 | 0.96497 | 0.99617 |
| Uterus Endometrium | 0.93939 | 1 | 0.96875 | 0.99384 | 0.9993 |
| Pancreas | **0.8333** | **0.8824** | **0.8571** | 0.90971 | 0.98499 |
| Soft Tissue | 0.97872 | 0.93878 | 0.95833 | 0.99084 | 0.99802 |
| Head Neck | 0.94444 | 0.97143 | 0.95775 | 0.9968 | 0.99958 |
| Brain | 0.94788 | 0.94788 | 0.94788 | 0.98363 | 0.99689 |
| micro avg | 0.93571 | 0.93552 | 0.9342 | 0.9765 | 0.99643 |
| macro avg | 0.94984 | 0.94788 | 0.94792 | 0.98536 | 0.99724 |
| weighted avg | **0.8571** | 0.94737 | 0.9 | 0.97051 | 0.99799 |

Table 1. Test performance for primary site origin prediction. The precision, recall, F1-score, mean average precision (mAP), and AUC-ROC are shown for each primary origin and aggregation for all classes by micro averaging, macro averaging, and weighted averaging.