**Title: Ulcerative Colitis Severity Classification using EfficientNetB4 Transfer Learning and Ordinal Categorical Cross-Entropy Loss Function**

**Model performance of all Mayo scores**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Measurement** | **Fold 1** | **Fold 2** | **Fold 3** | **Fold 4** | **Fold 5** | **Fold 6** | **Fold 7** | **Fold 8** | **Fold 9** | **Fold 10** | **Average** | **SD** | **Max** | **Min** |
| **Top-1 Accuracy** | 0.8025 | 0.7989 | 0.8001 | 0.8031 | 0.7942 | 0.8013 | 0.8055 | 0.8037 | 0.7895 | 0.8001 | 0.7999 | 0.0045 | 0.8055 | 0.7895 |
| **Top-2 Accuracy** | 0.9834 | 0.9816 | 0.9828 | 0.9870 | 0.9864 | 0.9840 | 0.9852 | 0.9846 | 0.9810 | 0.9810 | 0.9837 | 0.0020 | 0.9870 | 0.9810 |
| **QWK** | 0.8730 | 0.8751 | 0.8677 | 0.8777 | 0.8697 | 0.8754 | 0.8812 | 0.8761 | 0.8709 | 0.8686 | 0.8735 | 0.0041 | 0.8812 | 0.8677 |
| **MAE** | 0.2023 | 0.2064 | 0.2052 | 0.2011 | 0.2100 | 0.2028 | 0.1968 | 0.2017 | 0.2064 | 0.2046 | 0.2037 | 0.0034 | 0.2100 | 0.1968 |
| **Precision** | 0.7447 | 0.7327 | 0.7426 | 0.7359 | 0.7361 | 0.7448 | 0.7523 | 0.7411 | 0.7383 | 0.7386 | 0.7407 | 0.0054 | 0.7523 | 0.7327 |
| **Recall** | 0.7309 | 0.7444 | 0.7211 | 0.7404 | 0.7276 | 0.7432 | 0.7496 | 0.7411 | 0.7377 | 0.7082 | 0.7344 | 0.0119 | 0.7496 | 0.7082 |
| **F1 Score** | 0.7374 | 0.7371 | 0.7301 | 0.7378 | 0.7317 | 0.7438 | 0.7508 | 0.7400 | 0.7376 | 0.7209 | 0.7367 | 0.0076 | 0.7508 | 0.7209 |

**Model performance of remission classification**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Measurement** | **Fold 1** | **Fold 2** | **Fold 3** | **Fold 4** | **Fold 5** | **Fold 6** | **Fold 7** | **Fold 8** | **Fold 9** | **Fold 10** | **Average** | **SD** | **Max** | **Min** |
| **Top-1 Accuracy** | 0.9650 | 0.9620 | 0.9656 | 0.9597 | 0.9585 | 0.9626 | 0.9668 | 0.9626 | 0.9591 | 0.9638 | 0.9626 | 0.0027 | 0.9668 | 0.9585 |
| **Top-2 Accuracy** | 0.8843 | 0.8743 | 0.8885 | 0.8766 | 0.8873 | 0.8837 | 0.8754 | 0.8784 | 0.8832 | 0.8903 | 0.8822 | 0.0054 | 0.8903 | 0.8743 |
| **Kappa** | 0.8780 | 0.8699 | 0.8786 | 0.8607 | 0.8550 | 0.8707 | 0.8846 | 0.8697 | 0.8581 | 0.8721 | 0.8697 | 0.0090 | 0.8846 | 0.8550 |
| **MAE** | 0.0350 | 0.0380 | 0.0344 | 0.0403 | 0.0415 | 0.0374 | 0.0332 | 0.0374 | 0.0409 | 0.0362 | 0.0374 | 0.0027 | 0.0415 | 0.0332 |
| **Precision** | 0.9444 | 0.9327 | 0.9507 | 0.9315 | 0.9334 | 0.9371 | 0.9459 | 0.9402 | 0.9319 | 0.9480 | 0.9396 | 0.0069 | 0.9507 | 0.9315 |
| **Recall** | 0.9338 | 0.9373 | 0.9288 | 0.9292 | 0.9219 | 0.9336 | 0.9388 | 0.9297 | 0.9262 | 0.9251 | 0.9304 | 0.0051 | 0.9388 | 0.9219 |
| **F1 Score** | 0.9390 | 0.9350 | 0.9393 | 0.9303 | 0.9275 | 0.9354 | 0.9423 | 0.9349 | 0.9290 | 0.9361 | 0.9349 | 0.0045 | 0.9423 | 0.9275 |

**Model performance of all Mayo scores without Squeeze Excitation layer**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Measurement** | **Fold 1** | **Fold 2** | **Fold 3** | **Fold 4** | **Fold 5** | **Fold 6** | **Fold 7** | **Fold 8** | **Fold 9** | **Fold 10** | **Average** | **SD** | **Max** | **Min** |
| **Top-1 Accuracy** | 0.7989 | 0.7948 | 0.7955 | 0.8043 | 0.7989 | 0.7912 | 0.7977 | 0.7995 | 0.8013 | 0.8031 | 0.7985 | 0.0037 | 0.8043 | 0.7912 |
| **Top-2 Accuracy** | 0.9840 | 0.9834 | 0.9828 | 0.9828 | 0.9792 | 0.9775 | 0.9834 | 0.9822 | 0.9822 | 0.9816 | 0.9819 | 0.0019 | 0.9840 | 0.9775 |
| **QWK** | 0.8720 | 0.8675 | 0.8758 | 0.8782 | 0.8731 | 0.8638 | 0.8688 | 0.8711 | 0.8682 | 0.8758 | 0.8714 | 0.0042 | 0.8782 | 0.8638 |
| **MAE** | 0.2064 | 0.2094 | 0.2052 | 0.2011 | 0.2064 | 0.2165 | 0.2076 | 0.2046 | 0.2058 | 0.2028 | 0.2066 | 0.0040 | 0.2165 | 0.2011 |
| **Precision** | 0.7334 | 0.7372 | 0.7284 | 0.7413 | 0.7389 | 0.7266 | 0.7386 | 0.7400 | 0.7403 | 0.7367 | 0.7361 | 0.0048 | 0.7413 | 0.7266 |
| **Recall** | 0.7274 | 0.7107 | 0.7358 | 0.7496 | 0.7573 | 0.7305 | 0.7316 | 0.7276 | 0.7358 | 0.7510 | 0.7357 | 0.0130 | 0.7573 | 0.7107 |
| **F1 Score** | 0.7302 | 0.7229 | 0.7315 | 0.7477 | 0.7480 | 0.7283 | 0.7345 | 0.7321 | 0.7377 | 0.7427 | 0.7356 | 0.0079 | 0.7480 | 0.7229 |

**Model performance of all Mayo scores without Spatial Attention layer**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Measurement** | **Fold 1** | **Fold 2** | **Fold 3** | **Fold 4** | **Fold 5** | **Fold 6** | **Fold 7** | **Fold 8** | **Fold 9** | **Fold 10** | **Average** | **SD** | **Max** | **Min** |
| **Top-1 Accuracy** | 0.7983 | 0.7960 | 0.7995 | 0.8043 | 0.7930 | 0.8049 | 0.7977 | 0.7948 | 0.8025 | 0.7966 | 0.7988 | 0.0038 | 0.8049 | 0.7930 |
| **Top-2 Accuracy** | 0.9822 | 0.9816 | 0.9846 | 0.9816 | 0.9822 | 0.9840 | 0.9846 | 0.9816 | 0.9834 | 0.9816 | 0.9827 | 0.0012 | 0.9846 | 0.9816 |
| **QWK** | 0.8703 | 0.8667 | 0.8704 | 0.8758 | 0.8675 | 0.8769 | 0.8708 | 0.8655 | 0.8714 | 0.8690 | 0.8704 | 0.0035 | 0.8769 | 0.8655 |
| **MAE** | 0.2082 | 0.2094 | 0.2064 | 0.2017 | 0.2117 | 0.1987 | 0.2070 | 0.2106 | 0.2046 | 0.2076 | 0.2066 | 0.0038 | 0.2117 | 0.1987 |
| **Precision** | 0.7335 | 0.8381 | 0.7336 | 0.7302 | 0.7293 | 0.7452 | 0.7435 | 0.7355 | 0.7544 | 0.7424 | 0.7486 | 0.0307 | 0.8381 | 0.7293 |
| **Recall** | 0.7230 | 0.7208 | 0.7282 | 0.7314 | 0.7216 | 0.7361 | 0.7464 | 0.7212 | 0.7455 | 0.7334 | 0.7308 | 0.0092 | 0.7464 | 0.7208 |
| **F1 Score** | 0.7273 | 0.7286 | 0.7304 | 0.7291 | 0.7249 | 0.7401 | 0.7437 | 0.7273 | 0.7492 | 0.7372 | 0.7338 | 0.0078 | 0.7492 | 0.7249 |