|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| class | Designation | Disease | nImages | nClasses | Distribution | link |
| Binary | DS1 | Malaria | 27,558 | 2 | Balanced | [1] |
| DS2 | Covid | 800 | 2 | Balanced | [2] |
| DS3 | breast cancer | 7909 | 2 | Imbalanced | [3] |
| DS4 | skin cancer | 9605 | 2 | Imbalanced | [4] |
| Multiclass | DS5 | brain tumor | 3264 | 4 | Imbalanced | [5] |
| DS6 | lung and colon cancer | 25,000 | 5 | Balanced | [6] |
| DS7 | skin lesions | 10,000 | 7 | Imbalanced | [7] |
| DS8 | colorectal cancer | 5,000 | 8 | Balanced | [8] |

# References

1. Malaria Cell Images Dataset. https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria. Accessed 9 Jul 2024.

2. COVID-19 and Normal X-ray Dataset: Balanced. https://www.kaggle.com/datasets/mrtejas/covid-19-and-normal-x-ray-dataset-balanced. Accessed 9 Jul 2024.

3. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A Dataset for Breast Cancer Histopathological Image Classification. IEEE Trans Biomed Eng. 2016;63:1455–62.

4. Melanoma Skin Cancer Dataset of 10000 Images. https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images. Accessed 9 Jul 2024.

5. Brain Tumor Classification (MRI). https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri. Accessed 9 Jul 2024.

6. Borkowski AA, Bui MM, Thomas LB, Wilson CP, DeLand LA, Mastorides SM. Lung and Colon Cancer Histopathological Image Dataset (LC25000). 2019.

7. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions - ViDIR Dataverse. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T. Accessed 9 Jul 2024.

8. Colorectal Histology MNIST. https://www.kaggle.com/datasets/kmader/colorectal-histology-mnist. Accessed 9 Jul 2024.