

Mô tả dữ liệu

1 Thông tin chung

Bộ dữ liệu trên được lấy trên kaggle, có tên là Skin Cancer MNIST: HAM10000. Bộ dữ liệu trên bao gồm 10000 nghìn ảnh về 7 loại bệnh ung thư da. Đây là đường link dẫn đến bộ dữ liệu: <https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000>

Bộ dữ liệu kia bao gồm 2 folder ảnh, tổng số ảnh trong cả 2 folder là 10000 ảnh và 1 file csv. File csv mô tả thông tin của bộ ảnh với các nhãn. Các thuộc tính trong file csv bao gồm :

- lesion-id: nhãn đánh số thứ tự của ảnh
- image-id: nhãn đánh tên của ảnh
- dx: nhãn của 7 loại bệnh về da
- dx-type: các nhãn về việc các loại bệnh về da được xác định bằng hình thức nào
- age: độ tuổi của bệnh nhân
- sex: giới tính của bệnh nhân
- localization: vị trí phát hiện bệnh

Dưới đây là thông tin về các nhãn có trong file csv

- * Các nhãn về bệnh ung thư da: gồm có 7 loại bệnh
 - nv: Melanocytic nevi (nốt ruồi) gồm 6705 ảnh
 - mel: melanoma (Ung thư hắc tố da) gồm 1113 ảnh
 - bkl: benign keratosis-like lesions (Bệnh dày sừng tiết bã) gồm 1099 ảnh
 - bcc: basal cell carcinoma (Ung thư biểu mô tế bào đáy) gồm 514 ảnh
 - Akiec: actinic keratosis (Dày sừng ánh sáng) gồm 327 ảnh
 - vasc: vascular lesions (tổn thương mạch máu) gồm 142 ảnh
 - df: dermatofibroma (những khối u vô hại nằm trong da) gồm 115 ảnh
- * các nhãn về cách giám định bệnh: gồm có 4 mã

- histo:histopathology (mô bệnh học)
- follow-up: follow-up examination
- consensus: expert consensus (ý kiến chuyên gia)
- confocal: in-vivo confocal microscopy (quan sát trên kính hiển vi)

2 Chi tiết

Dưới đây là thông tin về các bức ảnh:

- Mỗi ảnh đều có kích thước 600x450 pixel
- Mỗi ảnh trong tập dữ liệu trên có kích cỡ từ 250KB đến 300KB