**Instruction Manual**

*A guide for image extraction, labeling, and upload for the ISIC 2024 Challenge dataset*

First off, thank you so much for contributing data for this project! With this iteration in the ISIC Grand Challenge series, we will explore a new direction: skin cancer detection using 3D total body photography (TBP). While all prior ISIC challenges have focused on dermoscopy imaging, this will be the first to use images of lesions detected and cropped from 3D TBP. A single TBP capture can provide hundreds of example lesions for training AI algorithms and, since all lesions on the patient are imaged, provides patient-level contextual information for ugly duckling approaches. By focusing on TBP crops for the ISIC challenge, we are facilitating the advancement of AI applications in TBP clinical imaging systems. By nature of the TBP crops, this effort also intends to bridge the AI knowledge gap between algorithms that require harder-to-acquire dermoscopic photography and lower quality/lower resolution clinical photos and cell phone images.

The ISIC challenge task will be binary classification: differentiate images of malignant lesions from benign-to-intermediate lesions. Participants will be judged using area under the receiver operating curve (AUC); winners will be those with the highest AUC. The competition will be hosted on Kaggle (<https://www.kaggle.com/>).

The lesion images are automatically derived from existing Vectra WB360 TBP captures. When Vectra processes the 92 camera images (46 bi-camera positions), for each source position the program generates an enhanced image by applying color and contrast correction (to achieve an approximately even background skin "tone" or color, regardless of global lighting differences) and some amount of "sharpening". Individual lesions are detected/tagged in one of two ways; (1) manual lesion tagging, which is often done to attribute other clinical and dermoscopic photos to the lesion; (2) automated lesion detection as part of Canfield’s Vectra lesion analyzer (Vectra LA) software. Both forms of “tagging” assign a 3D coordinate to each lesion, which then guides cropping of the full 3D image in an automated fashion. The result of an individual crop is a 15 mm by 15 mm field-of-view image of a single lesion.

This document is intended to help you prepare and upload your data contribution.

Take note of several accompanying files:

* [1] *ex\_pathology\_metadata.csv*
* [2] *ISIC2024Tool\_v1.0.0\_README.txt*
* [3] *s1\_VectraDBTool input.py.txt*
* [4] s*1\_VectraDBTool inpyt\_NoZeroPad.py.txt*
* [5] s*2\_tile selection.py.txt*
* [6] s*3\_metadata reformat.py.txt*

Remove the .txt extension from the filenames [3]-[6].

**Step 1. Identify patients**

Identify the set of patients you wish to include in this dataset and assemble as list of MRNs.

**Step 2. Pathology metadata**

The most important step in the process is creating your lesion-level pathology metadata table. It is also the most time-intensive step of the entire process, as all subsequent steps are mostly automated. This table should include exactly one row for every biopsied lesion from your set of patients. There are five required elements to include in your table.

1. **mrn** – include a column that contains the MRN (patient-identifier) for that biopsied lesion
2. **lesion\_tag** – include a column that contains the lesion tag number of the biopsied lesion (i.e., 1, 2, 3, 4, …, etc.).
3. **biopsy\_date** – include a column that contains the date of the biopsy. Required date format is %m/%d/%Y (i.e., 12/14/2023 is December 14th, 2023).
4. **diagnosis** – include a column that lists the histologic diagnosis. Values must conform to the list of diagnoses below in **Table 1**.
5. **mel\_class** – include a column that distinguishes melanoma in-situ from invasive melanoma, when applicable. Values must conform to the list of diagnosis below. When unknown or not applicable, leave blank.

Name the file **lesion.csv** and save the file locally. You can use [1] *ex\_pathology\_metadata.csv* as a template for proper formatting.

In addition, we encourage you to provide additional diagnostic attributes listed in Table 1. Although it is not required, this information enhances the quality of diagnostic labels in the dataset!

*Note if there are additional patients who did not have a biopsied lesion, you must include additional rows for those patients. For these patients, set lesion\_tag = 0 and biopsy\_date = 1/1/2024. Leave diagnosis fields blank!*

*Note if you include lesions whose diagnoses were confirmed with a method other than histopathology following a biopsy (such as “reflectance confocal microscopy” or “expert assessment”), include an additional column* ***diagnosis\_confirm\_type*** *that distinguishes “histopathologic” confirmed lesions from other methods of confirmation.*

**Table 1. Structured diagnostic variables.**

|  |  |  |
| --- | --- | --- |
| Variable | Name | Values |
| Diagnosis | diagnosis | |  | | --- | | acrochordon | | actinic keratosis | | adnexal tumor | | AIMP | | angiofibroma or fibrous papule | | angioma | | atypical melanocytic proliferation | | basal cell carcinoma | | cyst | | dermatofibroma | | lentigo NOS | | lichenoid keratosis | | melanoma | | melanoma metastasis | | neurofibroma | | nevus | | pigmented benign keratosis | | seborrheic keratosis | | solar lentigo | | squamous cell carcinoma | | verruca | |
| Nevus growth pattern | nevus\_growth\_pattern | |  | | --- | | compound | | intradermal | | junctional | | unknown or N/A | |
| Atypia (applicable to nevi) | atypia | |  | | --- | | mild | | mild/moderate | | moderate | | moderate/severe | | severe  unknown or N/A | |
| Lentiginous growth pattern (applicable to nevi) | mel\_lent\_pat | yes  no  unknown or N/A |
| Melanoma class | mel\_class | |  | | --- | | invasive melanoma | | melanoma in situ | | Unknown or N/A | |
| Melanoma mitotic index (#/mm^2) | mel\_mit | Numeric |
| Melanoma thickness (mm) | mel\_thick | Numeric |
| BCC subtype: nodular | bcc\_nod | |  | | --- | | Yes | | No | | unknown or N/A | |
| BCC subtype: superficial | bcc\_sup | |  | | --- | | Yes | | No | | unknown or N/A | |
| BCC subtype: micronodular | bcc\_micro | |  | | --- | | Yes | | No | | unknown or N/A | |
| BCC subtype: infiltrative | bcc\_inf | |  | | --- | | Yes | | No | | unknown or N/A | |
| SCC subtype | scc\_type | |  | | --- | | SCC in situ | | SCC in situ bowenoid | | SCC invasive | | SCC keratoacanthomatous | | unknown or N/A | |

**Step 3. Tile extraction**

You have identified the patients for this challenge. Now you will export the data from your Vectra system using the tile extraction tool.

**3.1.** First, follow the export tool installation instructions provided in [2] *ISIC2024Tool\_v1.0.0\_README.txt* (provided to you by Brian D’Alessando via email with download link).

Next, you will create the input for the extraction tool as a line-delimited series of all MRNs identified in Step 1. The python script [3] *s1\_VectraDBTool input.py* will create the input in the form of a .txt file.

*NOTE: if your MRNs do NOT contain leading zeroes or are NOT all a fixed length, use [4] s1\_VectraDBTool input\_NoZeroPad.py.*

**3.2.** Open the file in a text editor and enter your parameters:

* **folder** location of the list of MRNs.
* **file** within “folder” containing the MRNs. This should be **lesion.csv** (the pathology metadata file created in Step 2.).
* **patient\_column\_idx** is the index of the column containing the MRNs. If it is the first column of the table (Column A), then patient\_column\_idx = 0.

**3.3.** Run the python script and check that a file named *VectraDBTool\_input.txt* was created in “folder”.

**3.4.** To batch process WB360 captures, run the **VectraDBTool** (provided by Canfield), select the desired database, then click "DermX Utils > Export ISIC 2024 tool data" from the top menu bar of the application. A pop-up box will allow you to choose an output directory to write the tiling data. *Tip: Export to a local folder instead of a network drive so the next steps will process faster*. A checkbox option to not overwrite existing data is available. The default when checked is to skip exports already completed.

After clicking OK, the tool will then prompt for MRNs to process. Copy in the contents from *VectraDBTool\_input.txt* and run the tool.

Take a break and allow the tool time to process. It may require several hours to process several hundred patients.

**Step 4. Tile selection**

You will now sub-select which tiles to upload to the ISIC Archive.

**4.1**. Open the second python script: [5] *s2\_tile selection.py*.

**4.2**. Enter your parameters

* **folder\_labels:** the folder listed in Step 3, containing the metadata file created in Step 2.
* **file\_labels**: the metadata file created in Step 2 (likely named “lesion.csv”)
* **patient\_col\_idx**: index of column containing MRN
* **lesion\_col\_idx**: index of column containing lesion tag number
* **bxday\_col\_idx**: index of column containing biopsy date
* **folder\_tiles:** location of the exported tiles from step 3.4.
* **file\_tiles**: set to ‘patients.csv’
* **out\_dir:** specify a new local output directory

**4.3** Run the python script. Take a break and allow time for this to process. It may take several hours. Will print “Complete!” in the console when finished.

**4.4** Inside the new (parent) folder, *out\_dir*, there should be a (child) folder named **tags**. Open **tags** using file explore/finder. Manually delete any tile that does not depict a primary lesion (i.e. scars, bandages, surgical sites), that depicts PHI (i.e. tattoo), or is otherwise completely obscured by some other object (i.e. hair, clothing).

**Step 5. Metadata formatting**

Here you will clean up the metadata file created in Step 4.

**5.1.** Open [6] *s3\_metadata reformat.py*

**5.2.** Enter your parameters

* **tile\_dir**: the out\_dir from step 4.2.

**5.3.** Run the python script. It will finish more quickly than the prior scripts. Check that a new file named **metadata\_revised.csv** has been created to know it has run successfully.

**Step 6. Upload to the ISIC Archive**

**6.1.** Compress all tile images into one zip file. You can do this by selecting **tags** and all individual patient sub-folders, right-click, compress. Name this file **tiles.zip**.

**6.2.** Go to <https://www.isic-archive.com/> and click **Upload Images**.

**6.3.** Log in or create a new account (upper right corner of page).

**6.4.** Select an existing Contributor or Create**New Contributor** (bottom of page). Contributor is synonymous with *institution*. For example, images from Drs. Halpern’s and Rotemberg’s clinics are uploaded under *Memorial Sloan Kettering Cancer Center*.

**6.5.** Click **New Cohort** (bottom of the page) and fill out the template with a suitable name and description. Choose a creative commons license and list an attribution for how your images should be credited in citations.

**6.6.** Next **to Zip Files**, select **Add file** and upload **tiles.zip**.

**6.7.** Next to **Metadata Files**, select **Add file** and upload **metadata\_revised.csv.**

**6.8.** Email [kurtansn@mskcc.org](mailto:kurtansn@mskcc.org) or [support@isic-archive.com](mailto:support@isic-archive.com) to inform us that you are finished, and I will do a final review of your images before publishing to The Archive!

**Thank you!**