

Simplify medical image classification using Amazon SageMaker Canvas

by Ramakant Joshi, Dariush Azimi, Jake Wen, and Sonu Kumar Singh | on 04 OCT 2023 | in [Amazon SageMaker Canvas](#), [Artificial Intelligence](#), [Intermediate \(200\)](#), [Technical How-to](#) | [Permalink](#) | [Comments](#) | [Share](#)

Analyzing medical images plays a crucial role in diagnosing and treating diseases. The ability to automate this process using machine learning (ML) techniques allows healthcare professionals to more quickly diagnose certain cancers, coronary diseases, and ophthalmologic conditions. However, one of the key challenges faced by clinicians and researchers in this field is the time-consuming and complex nature of building ML models for image classification. Traditional methods require coding expertise and extensive knowledge of ML algorithms, which can be a barrier for many healthcare professionals.

To address this gap, we used [Amazon SageMaker Canvas](#), a visual tool that allows medical clinicians to build and deploy ML models without coding or specialized knowledge. This user-friendly approach eliminates the steep learning curve associated with ML, which frees up clinicians to focus on their patients.

Amazon SageMaker Canvas provides a drag-and-drop interface for creating ML models. Clinicians can select the data they want to use, specify the desired output, and then watch as it automatically builds and trains the model. Once the model is trained, it generates accurate predictions.

This approach is ideal for medical clinicians who want to use ML to improve their diagnosis and treatment decisions. With Amazon SageMaker Canvas, they can use the power of ML to help their patients, without needing to be an ML expert.

Medical image classification directly impacts patient outcomes and healthcare efficiency. Timely and accurate classification of medical images allows for early detection of diseases that aides in effective treatment planning and monitoring. Moreover, the democratization of ML through accessible interfaces like Amazon SageMaker Canvas, enables a broader range of healthcare professionals, including those without extensive technical backgrounds, to contribute to the field of medical image analysis. This inclusive approach fosters collaboration and knowledge sharing and ultimately leads to advancements in healthcare research and improved patient care.

In this post, we'll explore the capabilities of Amazon SageMaker Canvas in classifying medical images, discuss its benefits, and highlight real-world use cases that demonstrate its impact on medical diagnostics.

Use case

Skin cancer is a serious and potentially deadly disease, and the earlier it is detected, the better chance there is for successful treatment. Statistically, skin cancer (e.g. Basal and squamous cell carcinomas) is one of the most common cancer types and leads to hundreds of thousands of deaths [worldwide](#) each year. It manifests itself through the abnormal growth of skin cells.

However, early diagnosis drastically increases the chances of recovery. Moreover, it may render surgical, radiographic, or chemotherapeutic therapies unnecessary or lessen their overall usage, helping to reduce healthcare costs.

The process of diagnosing skin cancer starts with a procedure called a dermoscopy[1], which inspects the general shape, size, and color characteristics of skin lesions. Suspected lesions then undergo further sampling and histological tests for confirmation of the cancer cell type. Doctors use multiple methods to detect skin cancer, starting with visual detection. The American Center for the Study of Dermatology developed a guide for the possible shape of melanoma, which is called [ABCD](#) (asymmetry, border, color, diameter) and is used by doctors for initial screening of the disease. If a suspected skin lesion is found, then the doctor takes a biopsy of the visible lesion on the skin and examines it microscopically for a benign or malignant diagnosis and the type of skin cancer. Computer vision models can play a valuable role in helping to identify suspicious moles or lesions, which enables earlier and more accurate diagnosis.

Creating a cancer detection model is a multi-step process, as outlined below:

1. Gather a large dataset of images from healthy skin and skin with various types of cancerous or precancerous lesions. This dataset needs to be carefully curated to ensure accuracy and consistency.
2. Use computer vision techniques to preprocess the images and extract relevant to differentiate between healthy and cancerous skin.
3. Train an ML model on the preprocessed images, using a supervised learning approach to teach the model to distinguish between different skin types.
4. Evaluate the performance of the model using a variety of metrics, such as precision and recall, to ensure that it accurately identifies cancerous skin and minimizes false positives.
5. Integrate the model into a user-friendly tool that could be used by dermatologists and other healthcare professionals to aid in the detection and diagnosis of skin cancer.

Overall, the process of developing a skin cancer detection model from scratch typically requires significant resources and expertise. This is where Amazon SageMaker Canvas can help simplify the time and effort for steps 2 – 5.

Solution overview

To demonstrate the creation of a skin cancer computer vision model without writing any code, we use a dermoscopy skin cancer image dataset published by Harvard Dataverse. We use the dataset, which can be found at [HAM10000](#) and consists of 10,015 dermatoscopic images, to build a skin cancer classification model that predicts skin cancer classes. A few key points about the dataset:

- The dataset serves as a training set for academic ML purposes.
- It includes a representative collection of all important diagnostic categories in the realm of pigmented lesions.
- A few categories in the dataset are: Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc)
- More than 50% of the lesions in the dataset are confirmed through histopathology (histo).

- The ground truth for the rest of the cases is determined through follow-up examination (`follow_up`), expert consensus (consensus), or confirmation by *in vivo* confocal microscopy (confocal).
- The dataset includes lesions with multiple images, which can be tracked using the `lesion_id` column within the `HAM10000_metadata` file.

We showcase how to simplify image classification for multiple skin cancer categories without writing any code using Amazon SageMaker Canvas. Given an image of a skin lesion, SageMaker Canvas image classification automatically classifies an image into benign or possible cancer.

Prerequisites

- Access to an [AWS](#) account with permissions to create the resources described in the steps section.
- An AWS Identity and Access Management ([AWS IAM](#)) user with full permissions to use Amazon SageMaker.

Walkthrough

1. Set-up SageMaker domain

- a. Create an Amazon SageMaker domain using steps outlined [here](#).
- b. Download the [HAM10000](#) dataset.

2. Set-up datasets

- a. Create an Amazon Simple Storage Service ([Amazon S3](#)) bucket with a unique name, which is `image-classification-<ACCOUNT_ID>` where ACCOUNT_ID is your unique AWS AccountNumber.

Amazon S3 > Buckets > Create bucket

Create bucket [Info](#)

Buckets are containers for data stored in S3. [Learn more](#)

General configuration

Bucket name

`image-classification-<ACCOUNT_ID>`

Bucket name must be globally unique and must not contain spaces or uppercase letters. [See rules for bucket naming](#)

AWS Region

US East (N. Virginia) us-east-1 ▼

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Figure 1 Creating bucket

- b. In this bucket create two folders: `training-data` and `test-data`.

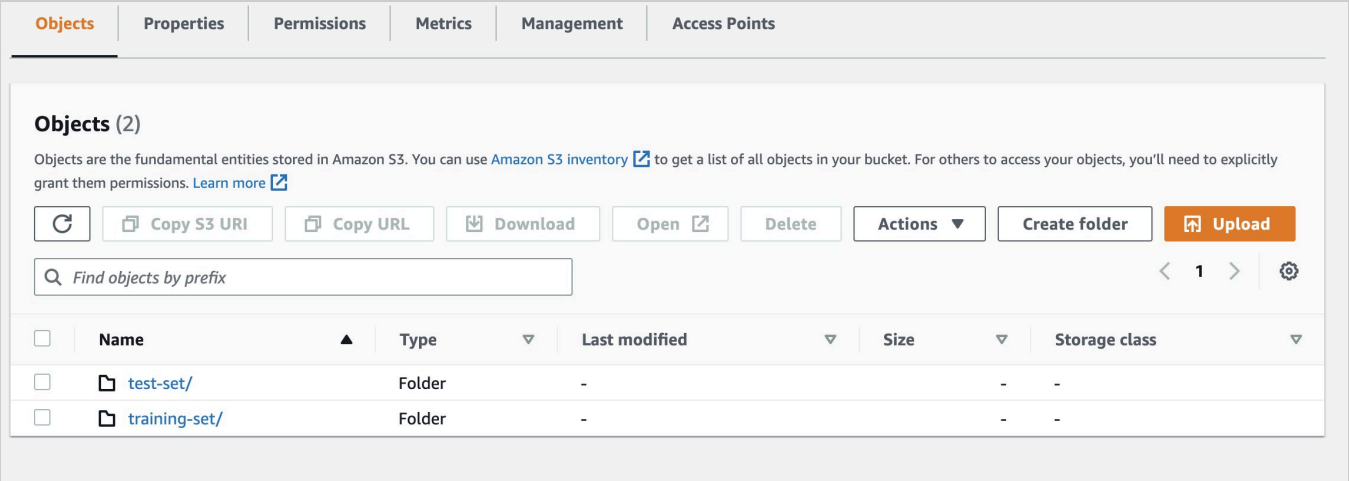


Figure 2 Create folders

- c. Under training-data, create seven folders for each of the skin cancer categories identified in the dataset: `akiec`, `bcc`, `bkl`, `df`, `mel`, `nv`, and `vasc`.

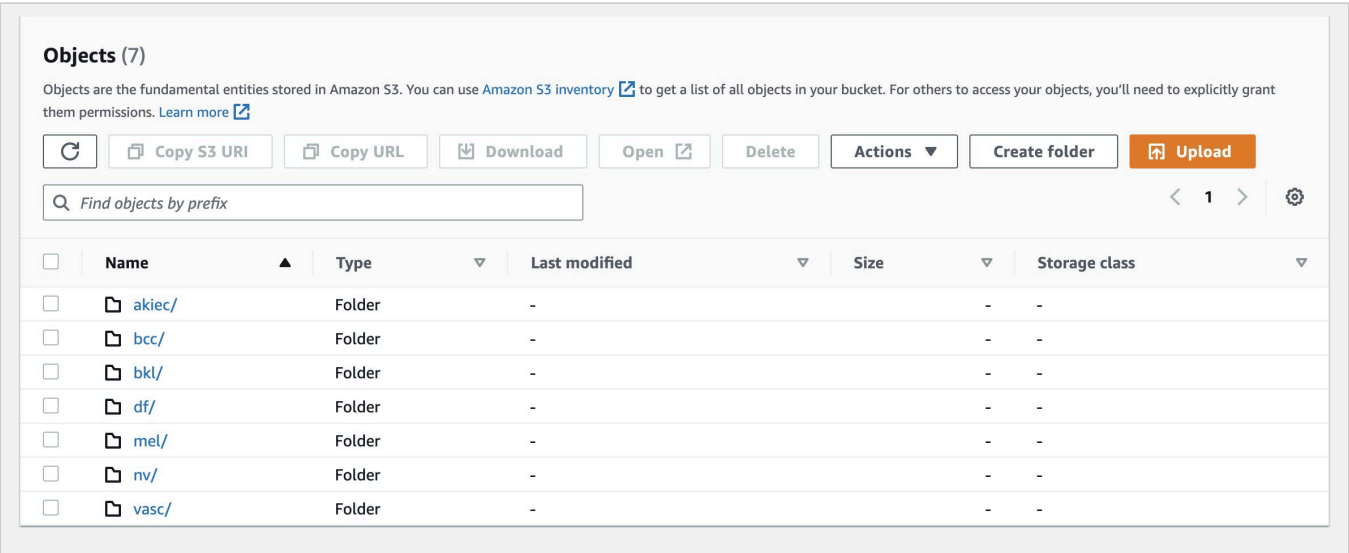


Figure 3 Folder View

- d. The dataset includes lesions with multiple images, which can be tracked by the `lesion_id-column` within the `HAM10000_metadata` file. Using the `lesion_id-column`, copy the corresponding images in the right folder (i.e., you may start with 100 images for each classification).

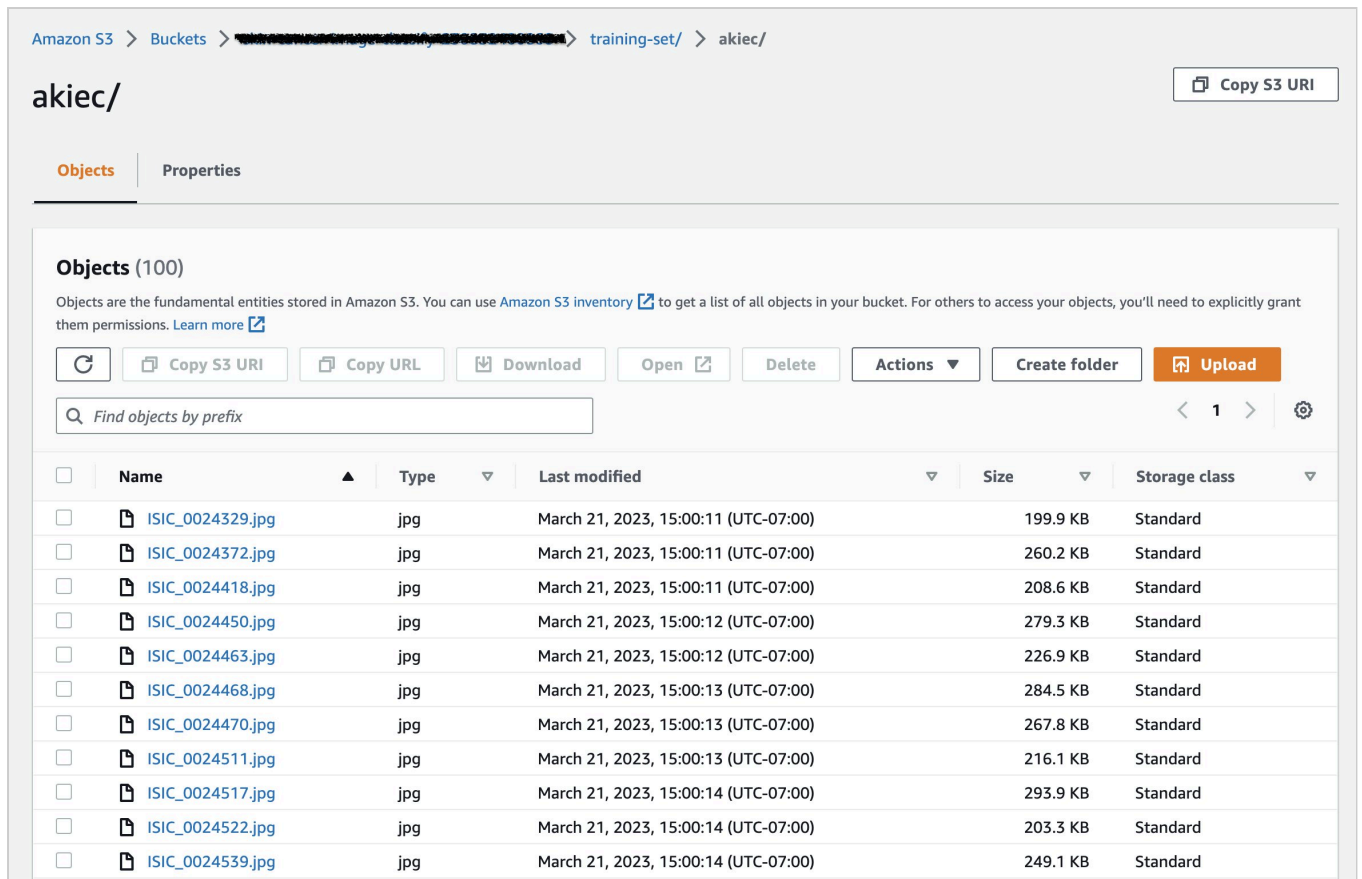


Figure 4 Listing Objects to import (Sample Images)

3. Use Amazon SageMaker Canvas

- Go to the **Amazon SageMaker** service in the console and select **Canvas** from the list. Once you are on the Canvas page, please select **Open Canvas** button.

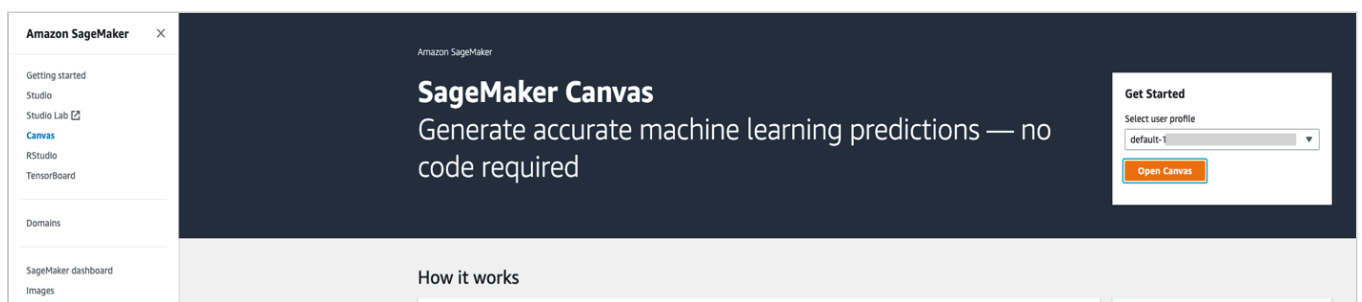


Figure 5 Navigate to Canvas

- Once you are on the Canvas page, select **My models** and then choose **New Model** on the right of your screen.

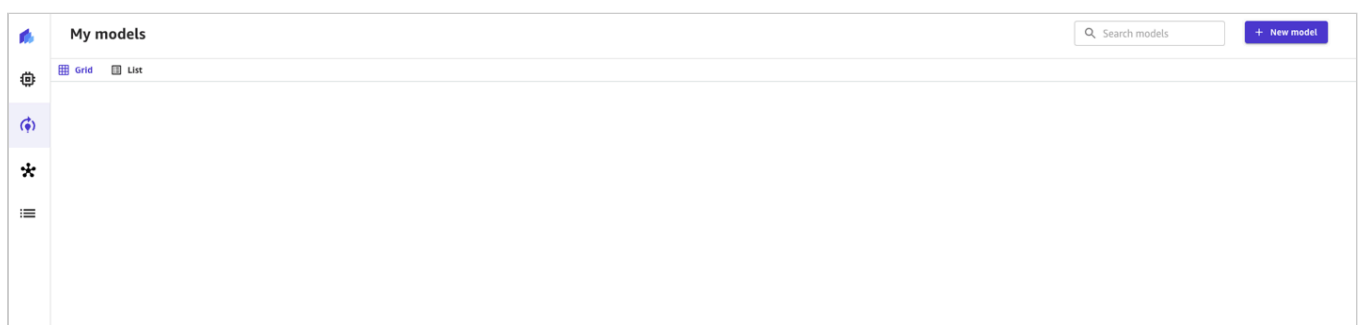


Figure 6 Creation of Model

- c. A new pop-up window opens up, where we name **image_classify** as the model's name and select Image analysis under the **Problem type**.

4. Import the dataset

- a. On the next page, please select **Create dataset** and in the pop-up box name the dataset as **image_classify** and select the **Create** button.

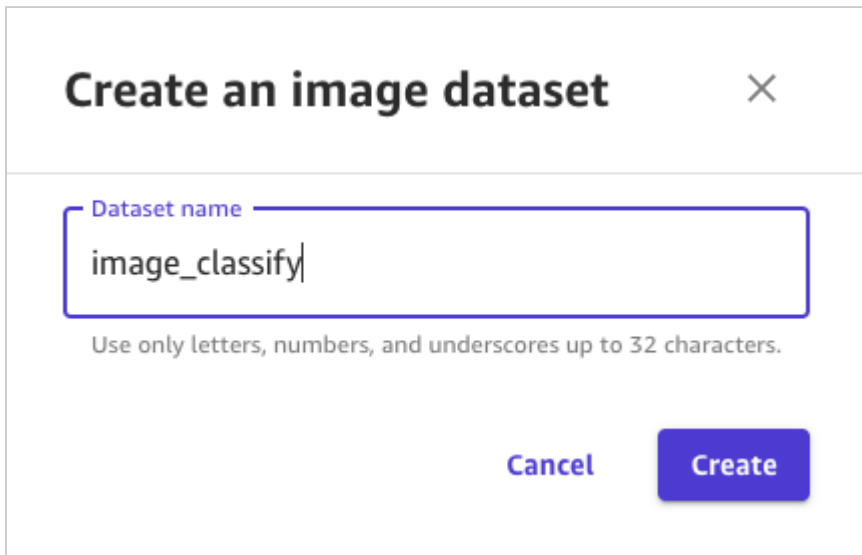


Figure 7 Creating dataset

- b. On the next page, change the **Data Source** to **Amazon S3**. You can also directly upload the images (i.e., **Local upload**).

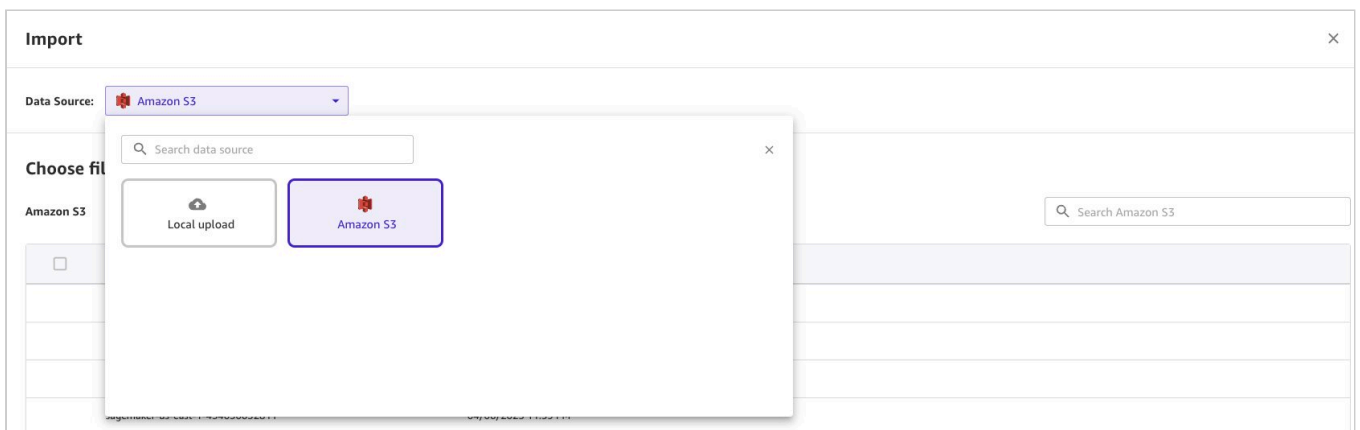


Figure 8 Import Dataset from S3 buckets

- c. When you select **Amazon S3**, you'll get the list of buckets present in your account. Select the parent bucket that holds the dataset into subfolder (e.g., **image-classify-2023** and select **Import data** button. This allows Amazon SageMaker Canvas to quickly label the images based on the folder names.
- d. Once, the dataset is successfully imported, you'll see the value in the Status column change to **Ready** from **Processing**.
- e. Now select your dataset by choosing **Select dataset** at the bottom of your page.

5. Build your model

- a. On the **Build** page, you should see your data imported and labelled as per the folder name in Amazon S3.

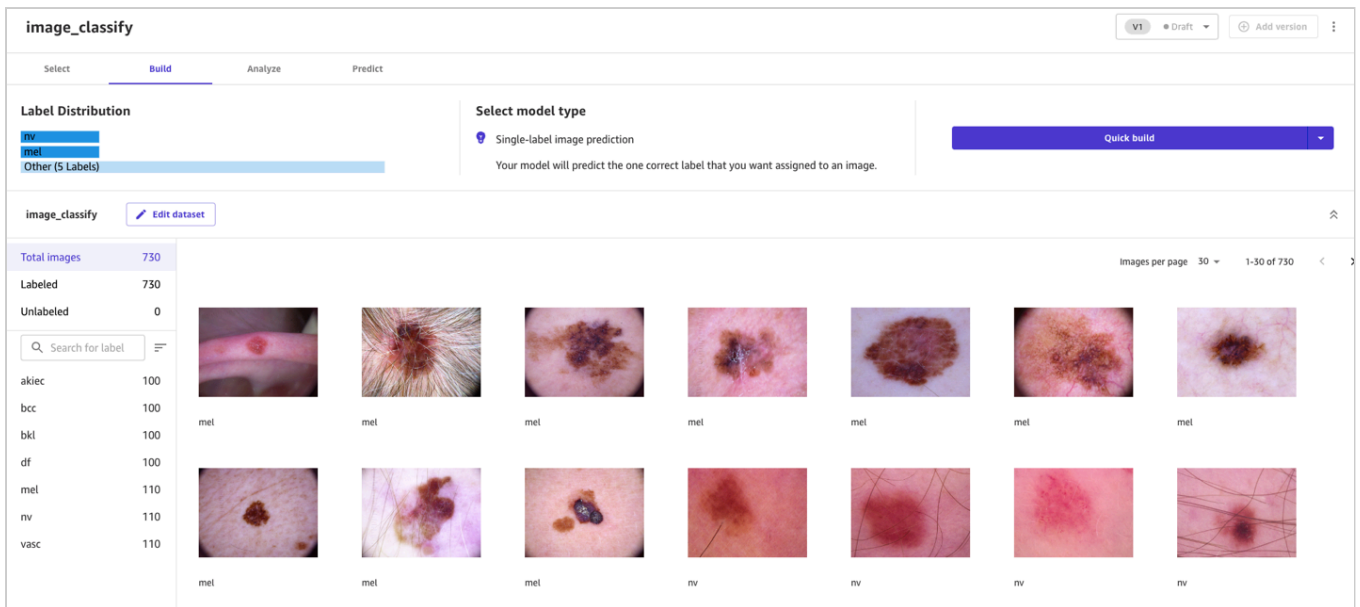


Figure 9 Labelling of Amazon S3 data

- b. Select the **Quick build** button (i.e., the red-highlighted content in the following image) and you'll see two options to build the model. First one is the **Quick build** and second one is **Standard build**. As name suggest quick build option provides speed over accuracy and it takes around 15 to 30 minutes to build the model. The standard build prioritizes accuracy over speed, with model building taking from 45 minutes to 4 hours to complete. Standard build runs experiments using different combinations of hyperparameters and generates many models in the backend (using SageMaker Autopilot functionality) and then picks the best model.
- c. Select **Standard build** to start building the model. It takes around 2–5 hours to complete.

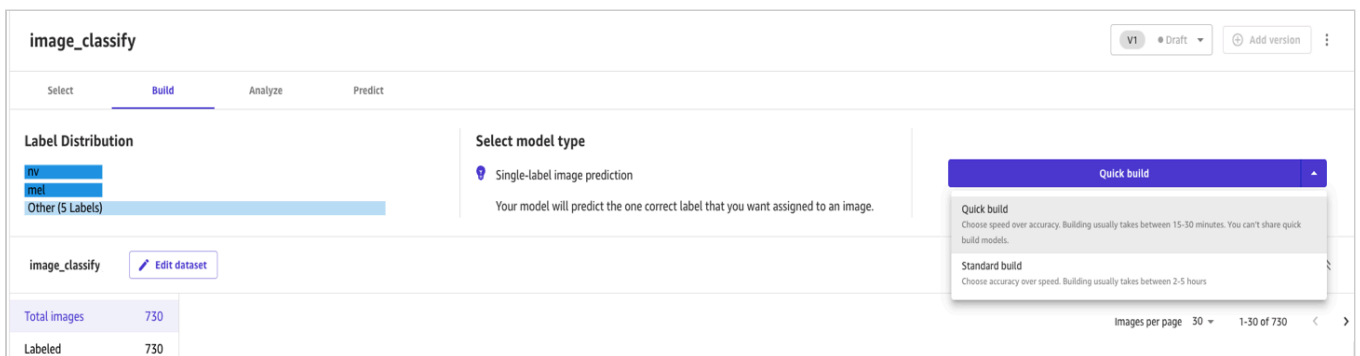


Figure 10 Doing Standard build

- d. Once model build is complete, you can see an estimated accuracy as shown in Figure 11.

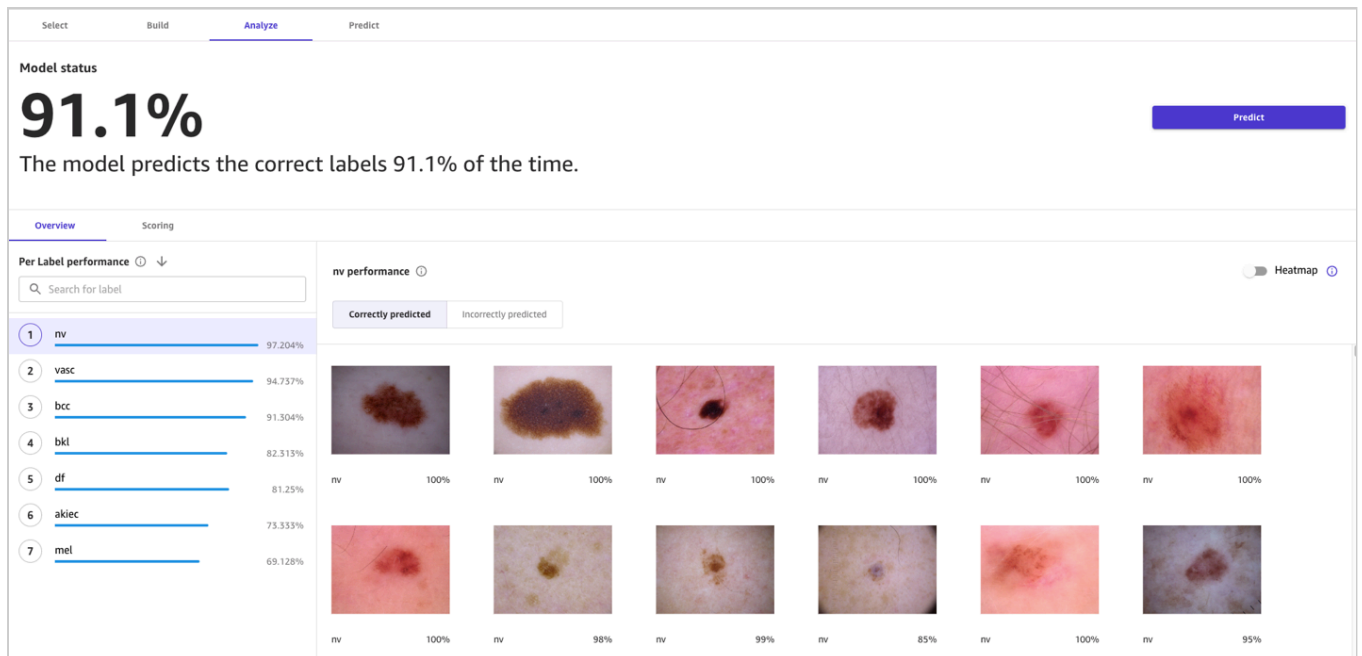


Figure 11 Model prediction

- e. If you select the **Scoring** tab, it should provide you insights into the model accuracy. Also, we can select the **Advanced metrics** button on the **Scoring** tab to view the precision, recall, and F1 score (A balanced measure of accuracy that takes class balance into account).
- f. The advanced metrics that Amazon SageMaker Canvas shows you depend on whether your model performs numeric, categorical, image, text, or time series forecasting predictions on your data. In this case, we believe recall is more important than precision because missing a cancer detection is far more dangerous than detecting correct. Categorical prediction, such as 2-category prediction or 3-category prediction, refers to the mathematical concept of classification. The [advanced metric](#) recall is the fraction of true positives (TP) out of all the actual positives (TP + false negatives). It measures the proportion of positive instances that were correctly predicted as positive by the model. Please refer this [A deep dive into Amazon SageMaker Canvas advanced metrics](#) for a deep dive on the advance metrics.



Figure 12 Advanced metrics

This completes the model creation step in Amazon SageMaker Canvas.

6. Test your model

- You can now choose the **Predict** button, which takes you to the **Predict** page, where you can upload your own images through **Single prediction** or **Batch prediction**. Please set the option of your choice and select **Import** to upload your image and test the model.

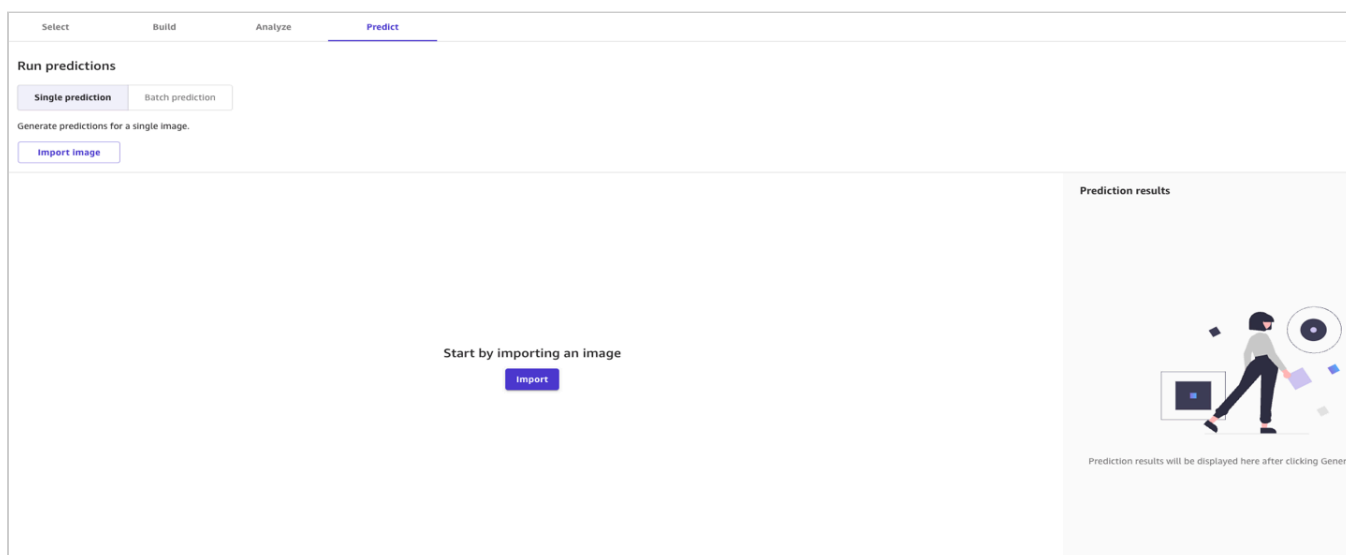


Figure 13 Test your own images

- b. Let's start by doing a single image prediction. Make sure you are on the **Single Prediction** and choose **Import image**. This takes you to a dialog box where you can choose to upload your image from **Amazon S3**, or do a **Local upload**. In our case, we select **Amazon S3** and browse to our directory where we have the test images and select any image. Then select **Import data**.

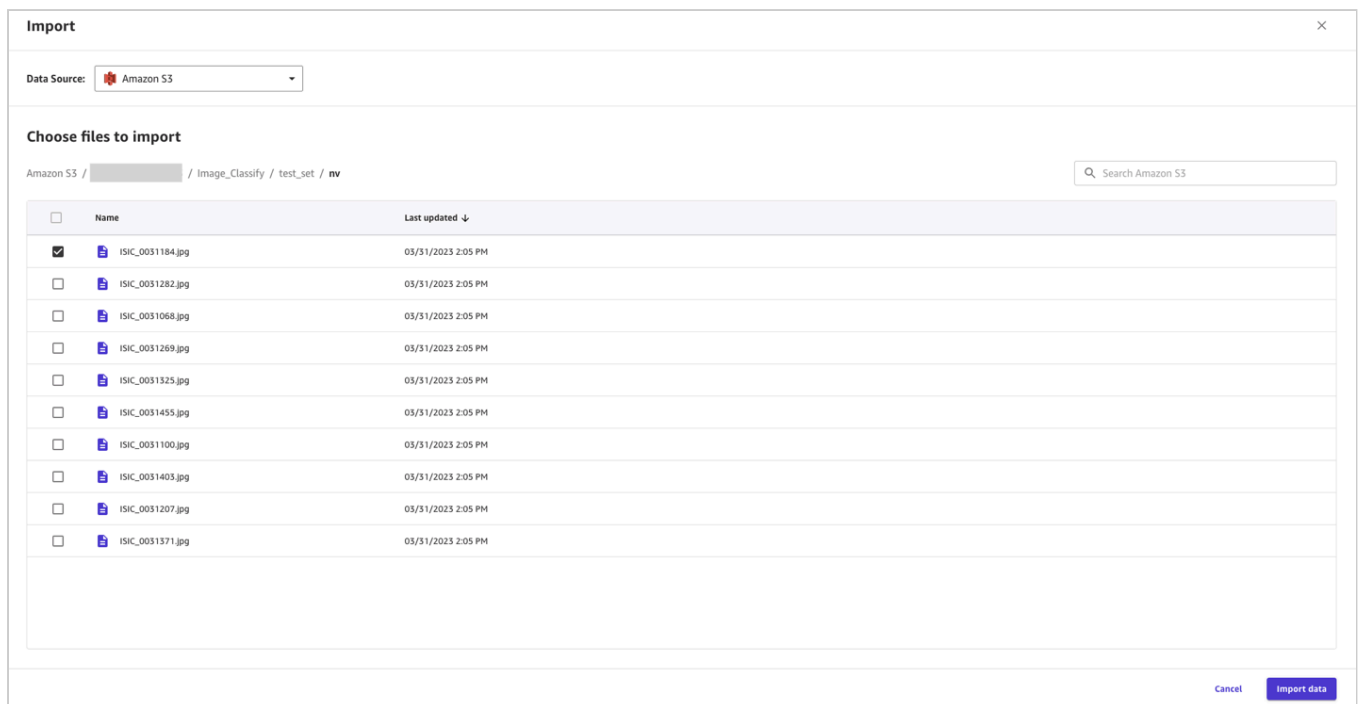


Figure 14 Single Image Prediction

- c. Once selected, you should see the screen says **Generating prediction results**. You should have your results in a few minutes as shown below.
- d. Now let's try the Batch prediction. Select **Batch prediction** under **Run predictions** and select the **Import new dataset** button and name it **BatchPrediction** and hit the **Create** button.

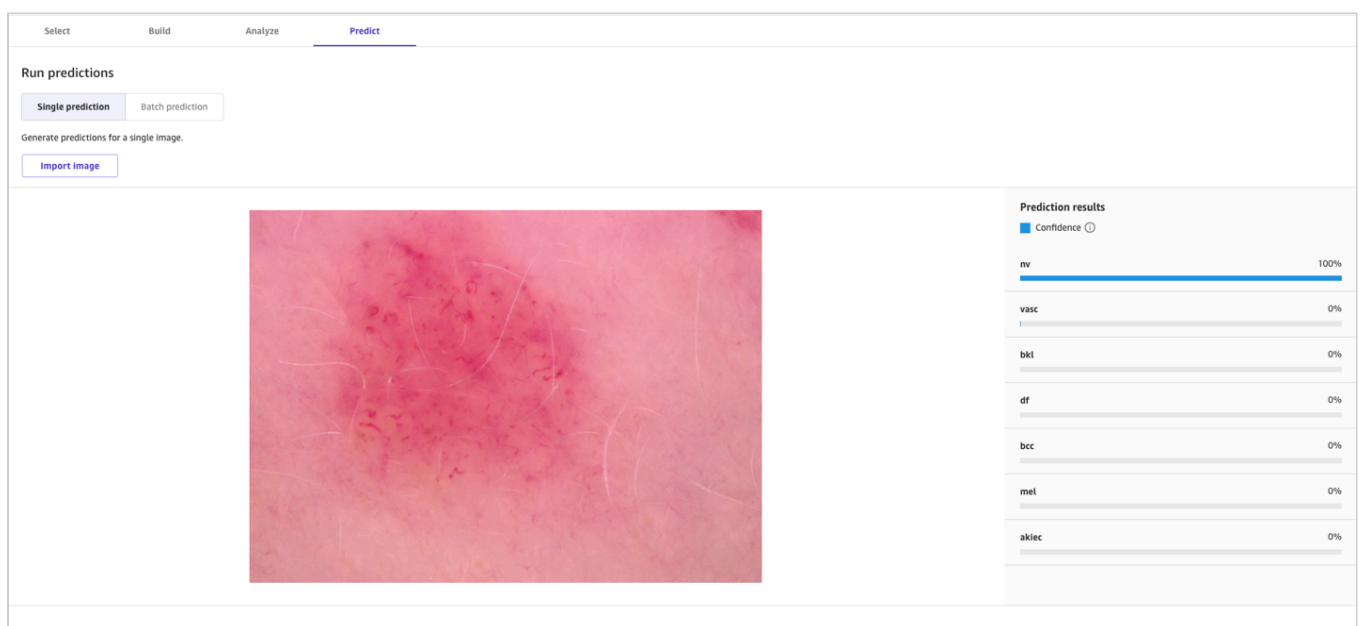


Figure 15 Single image prediction results

- e. On the next window, make sure you have selected Amazon S3 upload and browse to the directory where we have our test set and select the **Import data** button.

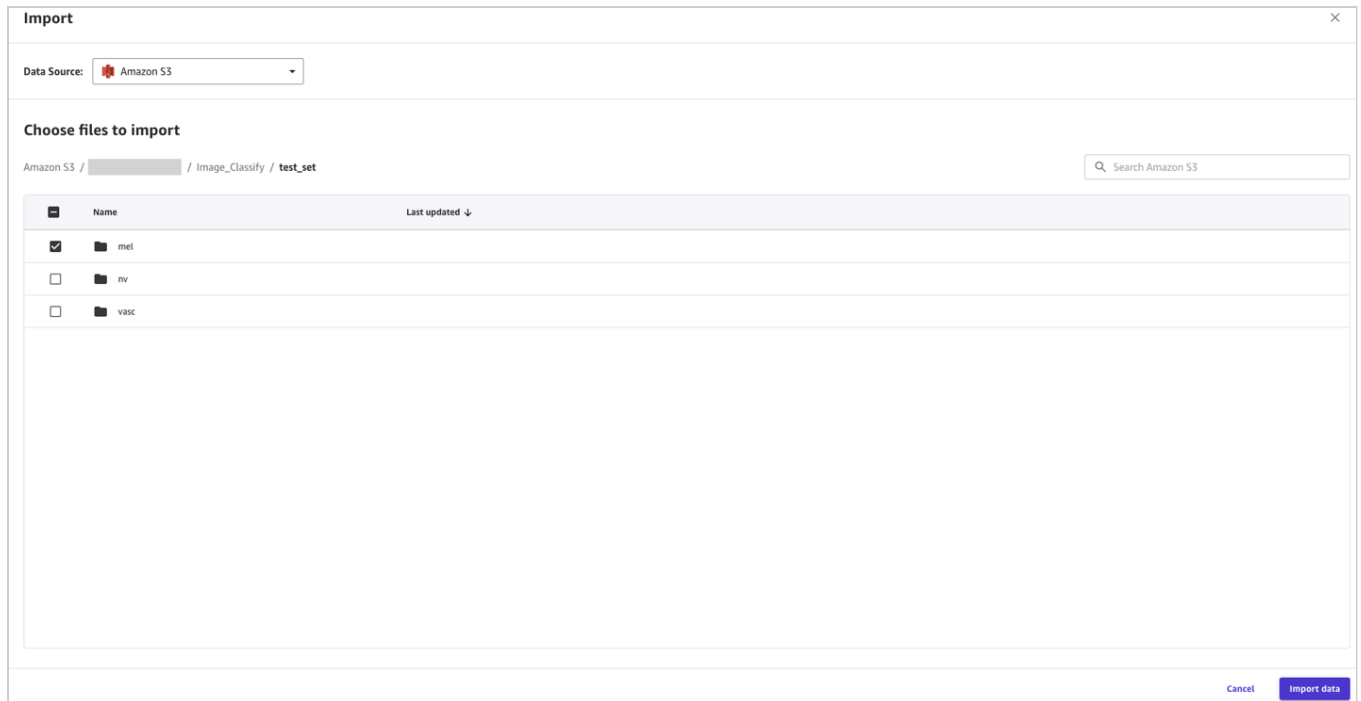


Figure 16 Batch Image Prediction

- f. Once the images are in **Ready** status, select the radio button for the created dataset and choose Generate predictions. Now, you should see the status of batch prediction batch to **Generating predictions**. Let's wait for few minutes for the results.
- g. Once the status is in **Ready** state, choose the dataset name that takes you to a page showing the detailed prediction on all our images.

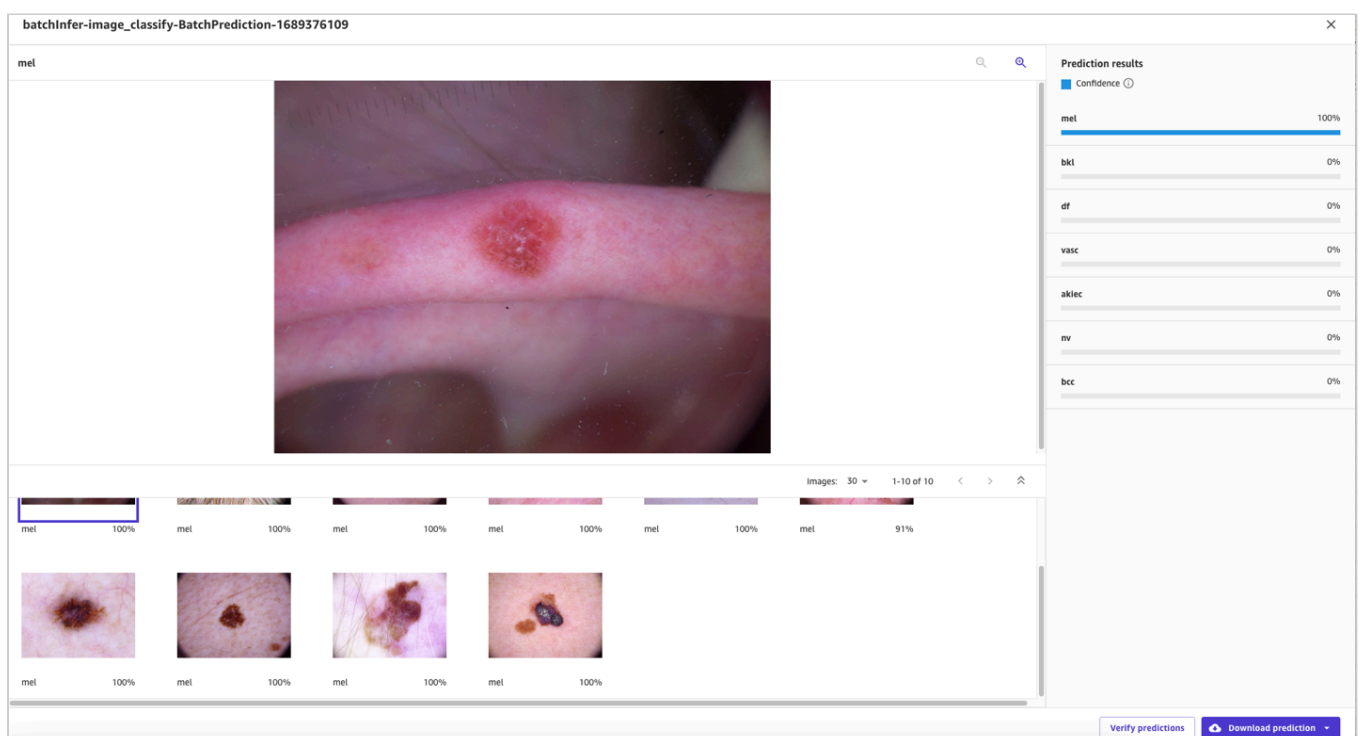


Figure 17 Batch image prediction results

- h. Another important feature of Batch Prediction is to be able to verify the results and also be able to download the prediction in a zip or csv file for further usage or sharing.

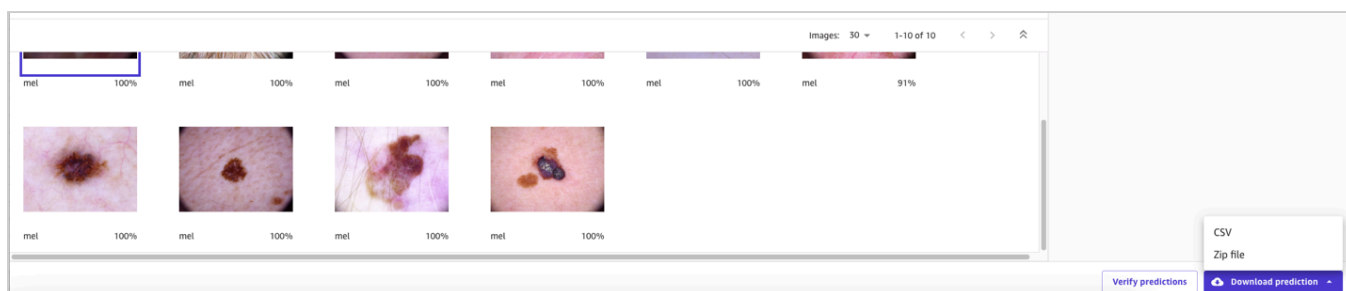


Figure 18 Download prediction

With this you have successfully been able to create a model, train it, and test its prediction with Amazon SageMaker Canvas.

Cleaning up

Choose **Log out** in the left navigation pane to log out of the Amazon SageMaker Canvas application to stop the consumption of [SageMaker Canvas workspace instance hours](#) and release all resources.

Citation

[1]Fraivan M, Faouri E. [On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning](#). Sensors (Basel). 2022 Jun 30;22(13):4963. doi: 10.3390/s22134963. PMID: 35808463; PMCID: PMC9269808.

Conclusion

In this post, we showed you how medical image analysis using ML techniques can expedite the diagnosis skin cancer, and its applicability to diagnosing other diseases. However, building ML models for image classification is often complex and time-consuming, requiring coding expertise and ML knowledge. Amazon SageMaker Canvas addressed this challenge by providing a visual interface that eliminates the need for coding or specialized ML skills. This empowers healthcare professionals to use ML without a steep learning curve, allowing them to focus on patient care.

The traditional process of developing a cancer detection model is cumbersome and time-consuming. It involves gathering a curated dataset, preprocessing images, training a ML model, evaluate its performance, and integrate it into a user-friendly tool for healthcare professionals. Amazon SageMaker Canvas simplified the steps from preprocessing to integration, which reduced the time and effort required for building a skin cancer detection model.

In this post, we delved into the powerful capabilities of Amazon SageMaker Canvas in classifying medical images, shedding light on its benefits and presenting real-world use cases that showcase its profound impact on medical diagnostics. One such compelling use case we explored was skin cancer detection and how early diagnosis often significantly enhances treatment outcomes and reduces healthcare costs.

It is important to acknowledge that the accuracy of the model can vary depending on factors, such as the size of the training dataset and the specific type of model employed. These variables play a role in determining the performance and reliability of the classification results.

Amazon SageMaker Canvas can serve as an invaluable tool that assists healthcare professionals in diagnosing diseases with greater accuracy and efficiency. However, it is vital to note that it isn't intended to replace the expertise and judgment of healthcare professionals. Rather, it empowers them by augmenting their capabilities and enabling more precise and expedient diagnoses. The human element remains essential in the decision-making process, and the collaboration between healthcare professionals and artificial intelligence (AI) tools, including Amazon SageMaker Canvas, is pivotal in providing optimal patient care.

About the authors



Ramakant Joshi is an AWS Solutions Architect, specializing in the analytics and serverless domain. He has a background in software development and hybrid architectures, and is passionate about helping customers modernize their cloud architecture.



Jake Wen is a Solutions Architect at AWS, driven by a passion for Machine Learning, Natural Language Processing, and Deep Learning. He assists Enterprise customers in achieving modernization and scalable deployment in the Cloud. Beyond the tech world, Jake finds delight in skateboarding, hiking, and piloting air drones.



Sonu Kumar Singh is an AWS Solutions Architect, with a specialization in analytics domain. He has been instrumental in catalyzing transformative shifts in organizations by enabling data-driven decision-making thereby fueling innovation and growth. He enjoys it when something he designed or created brings a positive impact. At AWS his intention is to help customers extract value out of AWS's 200+ cloud services and empower them in their cloud journey.



Dariush Azimi is a Solution Architect at AWS, with specialization in Machine Learning, Natural Language Processing (NLP), and microservices architecture with Kubernetes. His mission is to empower organizations to harness the full potential of their data through comprehensive end-to-end solutions encompassing data storage, accessibility, analysis, and predictive capabilities.

Comments

What do you think?

9 Responses



Awesome.
Love it!



Helpful



tl;dr

o Comments

Login ▼

G

Start the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS

Name



Share

Best Newest Oldest

Be the first to comment.

Subscribe

Privacy

Do Not Sell My Data