MedSigLIP model card

Page Summary

Model documentation: MedSigLIP (https://developers.google.com/health-ai-developer-foundations/medsiglip)

Resources:

- Model on Google Cloud Model Garden: MedSigLIP (https://console.cloud.google.com/vertex-ai/publishers/google/model-garden/medsiglip)
- Model on Hugging Face: MedSigLIP (https://huggingface.co/google/medsiglip-448)
- GitHub repository (supporting code, Colab notebooks, discussions, and issues): MedSigLIP (https://github.com/google-health/medsiglip)
- Quick start notebook: GitHub (https://github.com/google-health/medsiglip/blob/main/notebooks/quick_start_with_hugging_face.ipynb)
- Fine-tuning notebook: GitHub (https://github.com/google-health/medsiglip/blob/main/notebooks/fine_tune_with_hugging_face.ipynb)
- Support: See Contact (https://developers.google.com/health-ai-developer-foundations/medsiglip/get-started.md#contact)
- License: The use of MedSigLIP is governed by the Health AI Developer Foundations terms use (https://developers.google.com/health-ai-developer-foundations/terms).

Author: Google

Model information

This section describes the MedSigLIP model and how to use it.

Description

MedSigLIP is a variant of SigLIP (https://arxiv.org/abs/2303.15343) (Sigmoid Loss for Language Image Pre-training) that is trained to encode medical images and text into a common embedding space. Developers can use MedSigLIP to accelerate building healthcare-based AI applications. MedSigLIP contains a 400M parameter vision encoder and 400M parameter text encoder, it supports 448x448 image resolution with up to 64 text tokens.

MedSigLIP was trained on a variety of de-identified medical image and text pairs, including chest X-rays, dermatology images, ophthalmology images, histopathology slides, and slices of CT and MRI volumes, along with associated descriptions or reports. This training data was combined with natural (non-medical) image and text pairs to retain MedSigLIP's ability to parse natural images.

MedSigLIP is recommended for medical image interpretation applications without a need for text generation, such as data-efficient classification, zero-shot classification, and semantic image retrieval. For medical applications that require text generation, MedGemma (http://goo.gle/medgemma) is recommended.

How to use

Below are some example code snippets to help you quickly get started running the MedSigLIP model locally. If you want to use the model at scale, we recommend that you create a production version using Model Garden (https://console.cloud.google.com/vertex-ai/publishers/google/model-garden/medsiglip).

```
import numpy as np
from PIL import Image
import requests
from transformers import AutoProcessor, AutoModel
from tensorflow.image import resize as tf_resize
import torch

device = "cuda" if torch.cuda.is_available() else "cpu"

model = AutoModel.from_pretrained("google/medsiglip-448").to(device)
processor = AutoProcessor.from_pretrained("google/medsiglip-448")

# Download sample image
! wget -nc -q https://storage.googleapis.com/dx-scin-public-data/dataset/images/34450
! wget -nc -q https://storage.googleapis.com/dx-scin-public-data/dataset/images/-5669
imgs = [Image.open("3445096909671059178.png").convert("RGB"), Image.open("-5669089898


# If you want to reproduce the results from MedSigLIP evals, we recommend a
# resizing operation with `tf.image.resize` to match the implementation with the
# Big Vision library (https://github.com/google-research/big_vision/blob/0127fb6b337e
# Otherwise, you can rely on the Transformers image processor's built-in
# resizing (done automatically by default and uses `PIL.Image.resize`) or use
# another resizing method.
def resize(image):
    return Image.fromarray(
        tf_resize(
            images=image, size=[448, 448], method='bilinear', antialias=False
        ).numpy().astype(np.uint8)
    )


resized_imgs = [resize(img) for img in imgs]

texts = [
    "a photo of an arm with no rash",
    "a photo of an arm with a rash",
    "a photo of a leg with no rash",
    "a photo of a leg with a rash"
]

inputs = processor(text=texts, images=resized_imgs, padding="max_length", return_tens

with torch.no_grad():
    outputs = model(**inputs)

logits_per_image = outputs.logits_per_image
probs = torch.softmax(logits_per_image, dim=1)

for n_img, img in enumerate(imgs):
    display(img)  # Note this is an IPython function that will only work in a Ju
    for i, label in enumerate(texts):
        print(f"{probs[n_img][i]:.2%} that image is '{label}'")

# Get the image and text embeddings
print(f"image embeddings: {outputs.image_embeds}")
print(f"text embeddings: {outputs.text_embeds}")
```

✦**Code Tutor**

Examples

See the following Colab notebooks for examples of how to use MedSigLIP:

- To give the model a quick try, running it locally with weights from Hugging Face, see Quick start notebook in Colab

(https://colab.research.google.com/github/google-health/medsiglip/blob/main/notebooks/quick_start_with_hugging_face.ipynb)
.

- For an example of fine-tuning the model, see the Fine-tuning notebook in Colab (https://colab.research.google.com/github/google-health/medsiglip/blob/main/notebooks/fine_tune_with_hugging_face.ipynb)
.

Model architecture overview

MedSigLIP is based on SigLIP-400M (Zhai et al., 2023 (https://openaccess.thecvf.com/content/ICCV2023/html/Zhai_Sigmoid_Loss_for_Language_Image_Pre-Training_ICCV_2023_paper.html)
) and is the same encoder that powers image interpretation in the MedGemma (http://goo.gle/medgemma) generative model. MedSigLIP's image component is a 400M vision transformer and its text component is a 400M text transformer.

Technical specifications

- **Model type**: Two tower encoder architecture comprised of a vision transformer and text transformer
- **Image resolution**: 448 x 448
- **Context length**: 64 tokens
- **Modalities**: Image, text
- **Key publication**: https://arxiv.org/abs/2507.05201 (https://arxiv.org/abs/2507.05201)
- **Model created**: July 9, 2025
- **Model version**: 1.0.0

Can I help?

Citation

When using this model, please cite: Sellergren, Andrew, et al. "MedGemma Technical Report." *arXiv preprint arXiv:2507.05201* (2025).

```
@article{sellergren2025medgemma,
  title={MedGemma Technical Report},
  author={Sellergren, Andrew and Kazemzadeh, Sahar and Jaroensri, Tiam and Kiraly, At
  journal={arXiv preprint arXiv:2507.05201},
  year={2025}
}
```

Inputs and outputs

Input:

MedSigLIP accepts images and text as inputs.

- Images, normalized to values in the range (-1, 1) and to 448 x 448 resolution
- Text string, such as a caption or candidate classification label

Output:

- Image embedding if input image is provided
- Text embedding if input text is provided
- Similarity score between the image and text

Performance and validation

MedSigLIP was evaluated across a range of medical image modalities, focusing on chest X-ray, pathology, dermatology and ophthalmology.

Key performance metrics

The following table summarizes zero-shot AUCs for Chest X-Ray Findings with Med-SigLIP and ELIXR (Xu et al., 2023 (https://arxiv.org/abs/2308.01317)), based on CXR evaluation data from ELIXR. In all cases, 518 examples were used for 2-class classification. Note that MedSigLIP accepts inputs of size 448x448 while ELIXR accepts inputs of size 1280x1280.

Finding	Med-SigLIP Zero-Shot	ELIXR Zero-Shot*
Enlarged Cardiomediastinum	0.858	0.800
Cardiomegaly	0.904	0.891
Lung Opacity	0.931	0.888
Lung Lesion	0.822	0.747

Finding	Med-SigLIP Zero-Shot	ELIXR Zero-Shot*
Consolidation	0.880	0.875
Edema	0.891	0.880
Pneumonia	0.864	0.881
Atelectasis	0.836	0.754
Pneumothorax	0.862	0.800
Pleural Effusion	0.914	0.930
Pleural Other	0.650	0.729
Fracture	0.708	0.637
Support Devices	0.852	0.894
Average	**0.844**	**0.824**

*Prior reported results from (Xu et al., 2023 (https://arxiv.org/abs/2308.01317))

The following table summarizes AUCs for Dermatology, Ophthalmology, and Pathology Findings with Med-SigLIP compared to existing HAI-DEF embedding models (Derm Foundation and Path Foundation, goo.gle/hai-def (http://goo.gle/hai-def)). Note that MedSigLIP accepts inputs of size 448x448 while Derm Foundation accepts inputs of size 448x448 and Path Foundation accepts inputs of size 224x224.

Domain	Finding	Size	Num Classes	Med-SigLIP Zero-Shot	Med-SigLIP Linear Probe	HAI-DEF Li Probe*
Dermatology	Skin Conditions	161279		0.851	0.881	0.843
Ophthalmology	Diabetic Retinopathy	31615		0.759	0.857	N/A
Pathology	Invasive Breast Cancer	50003		0.933	0.930	0.943
	Breast NP	50003		0.721	0.727	0.758
	Breast TF	50003		0.780	0.790	0.832
	Cervical Dysplasia	50003		0.889	0.864	0.898
	Prostate Cancer Needles Core Biopsy	50004		0.892	0.886	0.915
	Radical Prostatectomy	50004		0.896	0.887	0.921

Can I help?

Domain	Finding	Size	Num Classes	Med-SigLIP Zero-Shot	Med-SigLIP Linear Probe	HAI-DEF Linear Probe*
	TCGA Study Types	5000	10	0.922	0.970	0.964
	Tissue Types	5000	16	0.930	0.972	0.947
Average				**0.870**	**0.878**	**0.897**

*HAI-DEF pathology results are based on prior reported results from Yang et al., 2024 (https://arxiv.org/abs/2405.03162).

Data card

Dataset overview

Training

MedSigLIP was trained on a variety of de-identified medical image and text pairs, including chest X-rays, dermatology images, ophthalmology images, histopathology slides, and slices of CT and MRI volumes, along with associated descriptions or reports. This training data was combined with natural (non-medical) image and text pairs to retain MedSigLIP's ability to parse natural images.

Evaluation

MedSigLIP has been evaluated on a comprehensive set of evaluation datasets on 23 tasks acro modalities and benchmarked against modality-specific HAI-DEF models from Google.

Source

MedSigLIP training utilized a combination of public and private datasets.

This model was trained on diverse public datasets including MIMIC-CXR (chest X-rays and reports), Slake-VQA, PAD-UFES-20 (skin lesion images and data), SCIN (dermatology images), TCGA (cancer genomics data), CAMELYON (lymph node histopathology images), PMC-OA (biomedical literature with images), and Mendeley Digital Knee X-Ray (knee X-rays).

Additionally, multiple diverse proprietary datasets were licensed and incorporated (described next).

Data ownership and documentation

- MIMIC-CXR (https://physionet.org/content/mimic-cxr/2.1.0/): MIT Laboratory for Computational Physiology and Beth Israel Deaconess Medical Center (BIDMC).
- Slake-VQA (https://www.med-vqa.com/slake/): The Hong Kong Polytechnic University (PolyU), with collaborators including West China Hospital of Sichuan University and Sichuan Academy of Medical Sciences / Sichuan Provincial People's Hospital.
- PAD-UFES-20 (https://pmc.ncbi.nlm.nih.gov/articles/PMC7479321/): Federal University of Espírito Santo (UFES), Brazil, through its Dermatological and Surgical Assistance Program (PAD).
- SCIN (https://github.com/google-research-datasets/scin): A collaboration between Google Health and Stanford Medicine.
- TCGA (https://portal.gdc.cancer.gov/) (The Cancer Genome Atlas): A joint effort of National Cancer Institute and National Human Genome Research Institute. Data from TCGA are available via the Genomic Data Commons (GDC)
- CAMELYON (https://camelyon17.grand-challenge.org/Data/): The data was collected from Radboud University Medical Center and University Medical Center Utrecht in the Netherlands.
- PMC-OA (PubMed Central Open Access Subset) (https://catalog.data.gov/dataset/pubmed-central-open-access-subset-pmc-oa): Maintained by the National Library of Medicine (NLM) and National Center for Biotechnology Information (NCBI), which are part of the NIH.
- MedQA (https://arxiv.org/pdf/2009.13081): This dataset was created by a team of researchers led by Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits
- Mendeley Digital Knee X-Ray (https://data.mendeley.com/datasets/t9ndx37v5h/1): This dataset from Rani Channamma University, and is hosted on Mendeley Data.

In addition to the public datasets listed above, MedSigLIP was also trained on de-identified, lice datasets or datasets collected internally at Google from consented participants.

- **Radiology dataset 1:** De-identified dataset of different CT and MRI studies across body parts from a US-based radiology outpatient diagnostic center network.
- **Ophthalmology dataset 1 (EyePACS):** De-identified dataset of fundus images from diabetic retinopathy screening.
- **Dermatology dataset 1:** De-identified dataset of teledermatology skin condition images (both clinical and dermatoscopic) from Colombia.
- **Dermatology dataset 2:** De-identified dataset of skin cancer images (both clinical and dermatoscopic) from Australia.

- **Dermatology dataset 3:** De-identified dataset of non-diseased skin images from an internal data collection effort.
- **Pathology dataset 1:** De-identified dataset of histopathology H&E whole slide images created in collaboration with an academic research hospital and biobank in Europe. Comprises de-identified colon, prostate, and lymph nodes.
- **Pathology dataset 2:** De-identified dataset of lung histopathology H&E and IHC whole slide images created by a commercial biobank in the United States.
- **Pathology dataset 3:** De-identified dataset of prostate and lymph node H&E and IHC histopathology whole slide images created by a contract research organization in the United States.
- **Pathology dataset 4:** De-identified dataset of histopathology whole slide images created in collaboration with a large, tertiary teaching hospital in the United States. Comprises a diverse set of tissue and stain types, predominantly H&E.

Data citation

- **MIMIC-CXR:** Johnson, A., Pollard, T., Mark, R., Berkowitz, S., & Horng, S. (2024). MIMIC-CXR Database (version 2.1.0). PhysioNet. https://physionet.org/content/mimic-cxr/2.1.0/ *and* Johnson, Alistair E. W., Tom J. Pollard, Seth J. Berkowitz, Nathaniel R. Greenbaum, Matthew P. Lungren, Chih-Ying Deng, Roger G. Mark, and Steven Horng. 2019. "MIMIC-CXR, a de-Identified Publicly Available Database of Chest Radiographs with Free-Text Reports." *Scientific Data* 6 (1): 1–8.
- **SLAKE:** Liu, Bo, Li-Ming Zhan, Li Xu, Lin Ma, Yan Yang, and Xiao-Ming Wu. 2021.SLAKE: A Semantically-Labeled Knowledge-Enhanced Dataset for Medical Visual Question Answeri http://arxiv.org/abs/2102.09542.
- **PAD-UEFS-20:** Pacheco, Andre GC, et al. "PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones." Data in brief 32 (2020): 106221.
- **SCIN:** Ward, Abbi, Jimmy Li, Julie Wang, Sriram Lakshminarasimhan, Ashley Carrick, Bilson Campana, Jay Hartford, et al. 2024. "Creating an Empirical Dermatology Dataset Through Crowdsourcing With Web Search Advertisements." *JAMA Network Open 7* (11): e2446615–e2446615.
- **TCGA:** The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.
- **CAMELYON16:** Ehteshami Bejnordi, Babak, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A. W. M. van der Laak, et al. 2017.

"Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer." *JAMA 318* (22): 2199–2210.

- **Mendeley Digital Knee X-Ray:** Gornale, Shivanand; Patravali, Pooja (2020), "Digital Knee X-ray Images", Mendeley Data, V1, doi: 10.17632/t9ndx37v5h.1

De-identification/anonymization:

Google and its partners utilize datasets that have been rigorously anonymized or de-identified to ensure the protection of individual research participants and patient privacy.

Implementation information

Details about the model internals.

Software

Training was done using JAX (https://github.com/jax-ml/jax).

JAX allows researchers to take advantage of the latest generation of hardware, including TPUs, for faster and more efficient training of large models.

Use and limitations

Intended use

MedSigLIP is a machine learning-based software development tool that generates numerical representations from input images and associated text. These representations are referred to as embeddings. MedSigLIP is designed for use by software developers and researchers to facilitate the creation and development of third-party healthcare applications that involve medical images and text. MedSigLIP itself does not provide any medical functionality, nor is it intended to process or interpret medical data for a medical purpose. MedSigLIP is a software development tool and is not a finished product. Developers are responsible for training, adapting, and making meaningful changes to MedSigLip to accomplish their specific intended use.

The embeddings that MedSigLIP generates can be used for downstream tasks such as classification, regression, and semantic search. Numerical scores based on calculations performed on the embeddings can be thresholded for classification, or semantic search use-cases, allowing developers to control for precision and recall. Embedding-based models enable developers to create solutions that can be more compute efficient for fine-tuning classification tasks, such as training classifiers.. Thus, MedSigLIP is recommended for applications requiring strong classification performance without the need for text generation. MedSigLIP has been specifically pre-trained on a variety of de-identified pairs of medical images and text, including chest X-rays, CT slices, MRI slices, dermatology images, ophthalmology images, and histopathology patches. MedSigLip is intended to be used by software developers, to be adapted for use in image based applications in healthcare domains such as radiology, pathology, ophthalmology, and dermatology.

Benefits

- Provides strong baseline medical image and text encodings.
- Lightweight model that can be used in settings with limited high-bandwidth memory accelerator access.
- MedSigLIP's strong performance makes it efficient to adapt for downstream healthcare-based use cases, compared to models of similar size without medical data pre-training.

Limitations

MedSigLIP is not intended to be used without appropriate validation, adaptation, and/or making meaningful modification by developers for their specific use case. Without the above, outputs generated by the MedSigLip model are not intended to directly inform clinical diagnosis, patient management decisions, treatment recommendations, or any other direct clinical practice applications. Any software application developed using MedSigLip that is intended for a medica purpose must be independently validated and is subject to its own regulatory requirements.

When adapting MedSigLIP developer should consider the following:

- **Bias in validation data:** As with any research, developers should ensure that any downstream application is validated to understand performance using data that is appropriately representative of the intended use setting for the specific application (e.g., age, sex, gender, condition, imaging device, etc).
- **Data contamination concerns**: When evaluating the generalization capabilities of a model like MedSigLIP in a medical context, there is a risk of data contamination, where the model might have inadvertently seen related medical information during its pre-training, potentially overestimating its true ability to generalize to novel medical concepts. Developers should validate MedSigLIP on datasets not publicly available or otherwise made available to non-institutional researchers to mitigate this risk.

Last updated 2025-07-09 UTC.

Can I help?