The *COFE* Ecosystem

GaNDLF

Generally Nuanced Deep Learning Framework

gandlf.org

ML Commons

S. Pati, et al., *Nature Communications Engineering*, 2(23), 2023

OpenVINO

Model optimization for inference on low-resource environments

openvino.ai

intel

A. Demidovskij, et al.; *ICCV Workshop*, 783-787, 2019

HF Hub

Model deployment across multiple platforms & ecosystems

hf.co

S.M. Jain, *Introduction to Transformers for NLP*, 51-67, Berkeley, 2022

OpenFL

Federated Learning Library

openfl.io

P. Foley, et al., *Phys Med Biol (ITCR Special Issue)*, 67(21), 214001, 2022

MedPerf

Governance & Orchestration

medperf.org

ML Commons

A. Karargyris, et al.; *Nature Machine Intelligence* 5:799-810, 2023

Introduction to Model Design using *GaNDLF*

Sarthak Pati

Division of Computational Pathology
Department of Pathology & Laboratory Medicine
Indiana university School of Medicine

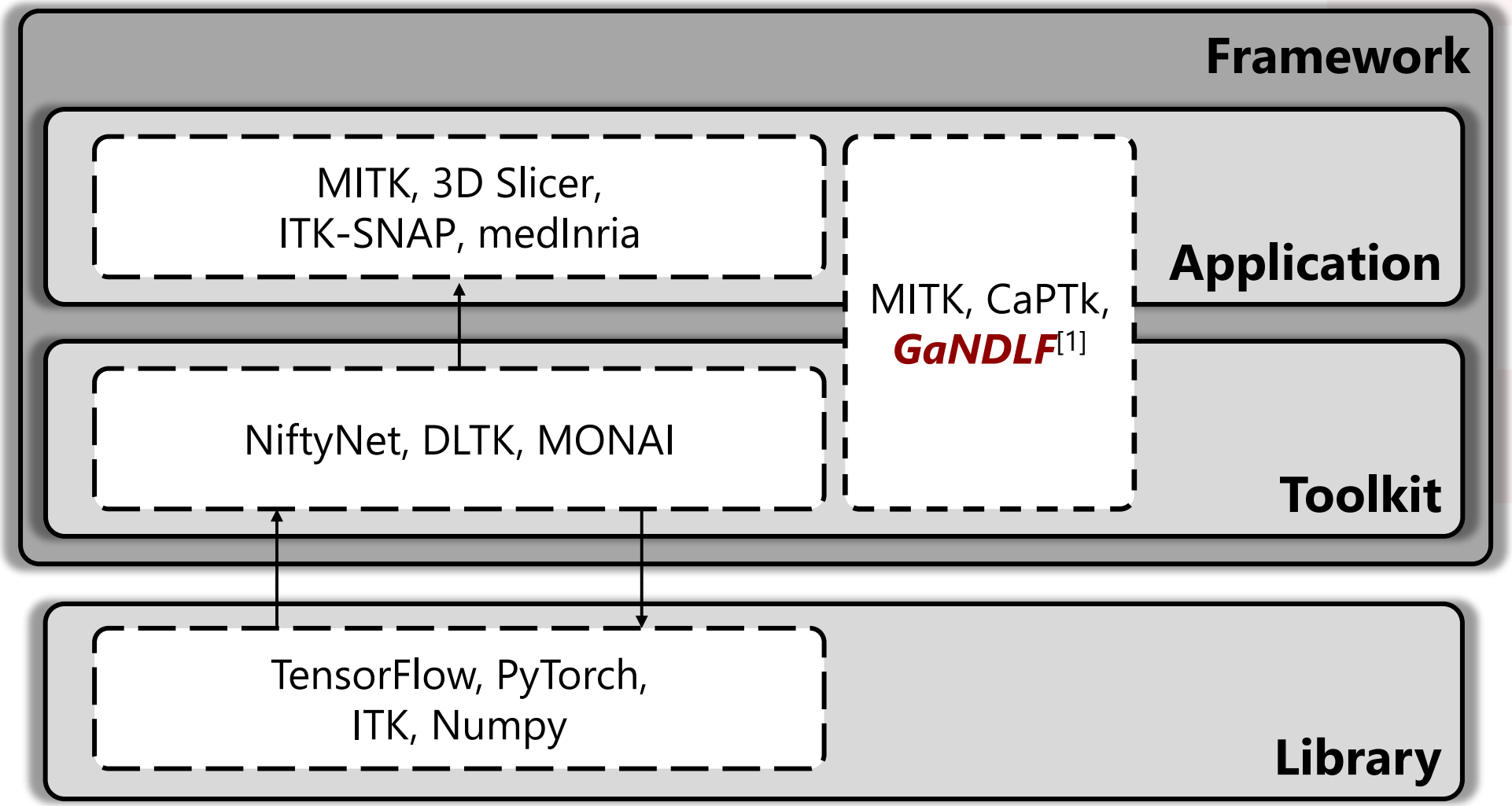
Medical Accuracy Working Group
MLCommons

@SarthakPati
patis@iu.edu

Table of Contents

Motivation

Software Categorization/Terminology



Straddles the line between toolkit and application

Focus on the end-user, with powerful user interfaces

Abstraction to libraries and general-level functionalities

Provides access to low-level machine functionality

Related Prior Work

- NiftyNet [1]
- DLTK [2]
- ANTsPyNet [3]
- DeepNeuro [4]

TensorFlow – limited deployment capability in clinical settings

- CaPTk [5]
- 3D-Slicer [6]
- MITK [7]

Excellent generalization but inferior performance compared to DL

- Specialized Algorithms
  - DeepMedic [8]
  - nnU-Net [9]
  - *Many more…*

Excellent performance but no generalization to new domains

[1] Gibson et al.; DOI:10.1016/j.cmpb.2018.01.025
[2] Pawlowski et al.; arXiv:1711.06853
[3] Tutison et al.; DOI:10.1038/s41598-021-87564-6
[4] Beers et al.; DOI:10.1007/s12021-020-09477-5
[5] Davatzikos et al.; DOI:10.1117/1.JMI.5.1.011018
[6] Pieper et al.; DOI:10.1109/ISBI.2004.1398617
[7] Wolf et al.; DOI:10.1117/12.535112
[8] Kamnitsas et al.; DOI:10.1016/j.media.2016.10.004
[9] Isensee et al.; DOI:10.1038/s41592-020-01008-z

Related Prior Work – PyTorch-specific

pymia[1] **InnerEye**[2] **fastai**[3] **MONAI**[4]

- Developed in PyTorch – exhaustive deployment capabilities for the clinical
- Developer-focused tools requiring experience in DL
- Not easy for DL developers to write their architectures in a generalizable way to span across domains
- Requires experience to write training pipelines for different problem domains
- Onus of training robust models on user
- Lack a single end-to-end application programming interface (API) for training/inference that can span across various problem domains.

[1] Jungo et al.; DOI:10.1016/j.cmpb.2020.105796
[2] Oktay et al.; DOI:10.1001/jamanetworkopen.2020.27426
[3] Howard J.; DOI:10.3390/info11020108
[4] Cardoso et al.; DOI:10.48550/arXiv.2211.02701

Steps of Research Project Lifecycle…

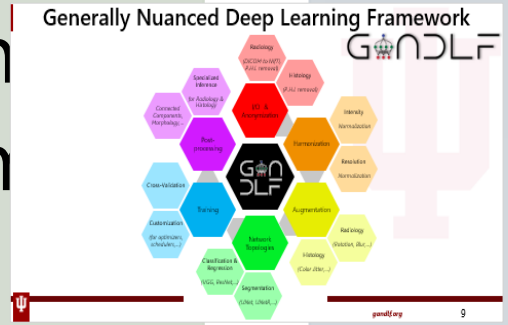
Conception & Design

- Use case
- Available data
- Potential Impact
- Experimental Design

The thought process

Development

- Data I/O
- Curation
- Preprocessin
- DL Algorithn



Evaluation

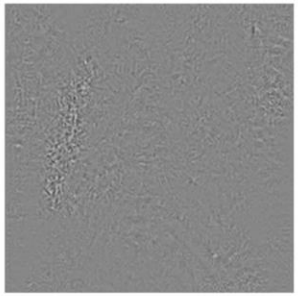
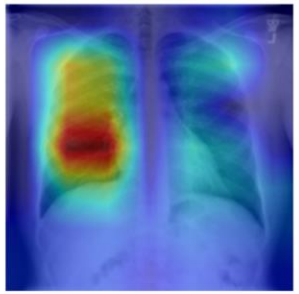
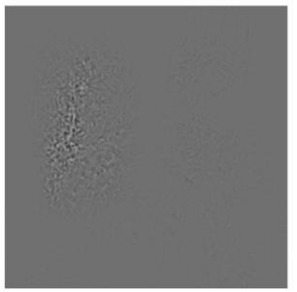
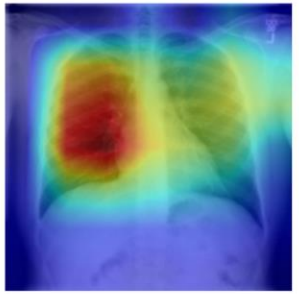
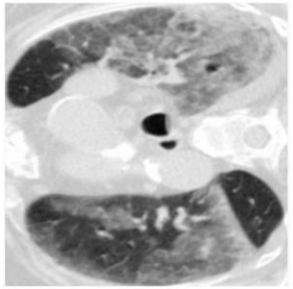
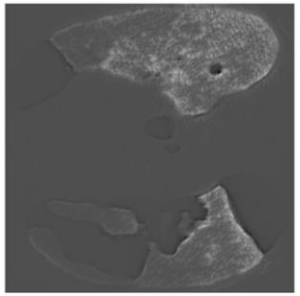
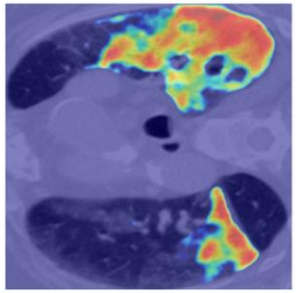
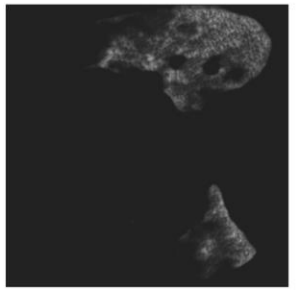
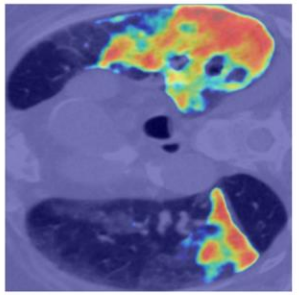
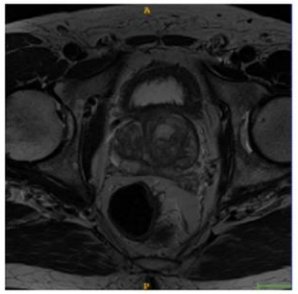
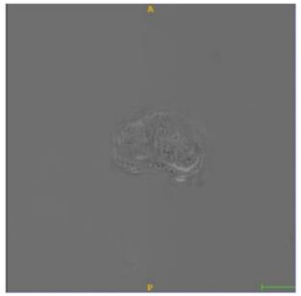
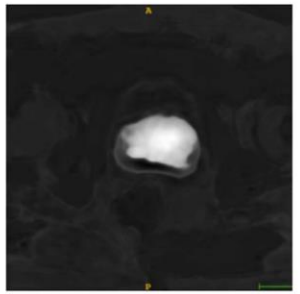
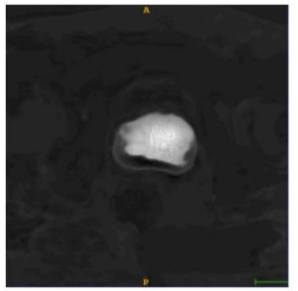
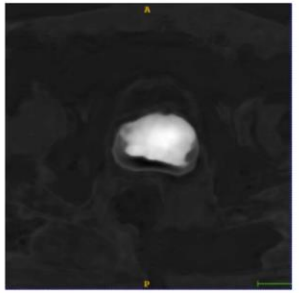
- Cross-validation
- Metrics

Reproducibility & Potential Translation

Generally Nuanced Deep Learning Framework

Current Features

Current Results: X-AI or Interpretability

	Image	Guided backpropagation	Grad-Cam	Guided Grad-Cam	Grad-Cam++
#1 Classification (2D)					
#2 Segmentation (2D)					
#3 Segmentation (3D)					

Gotkowski, Karol, et al. arXiv preprint arXiv:2007.00453 (2020).

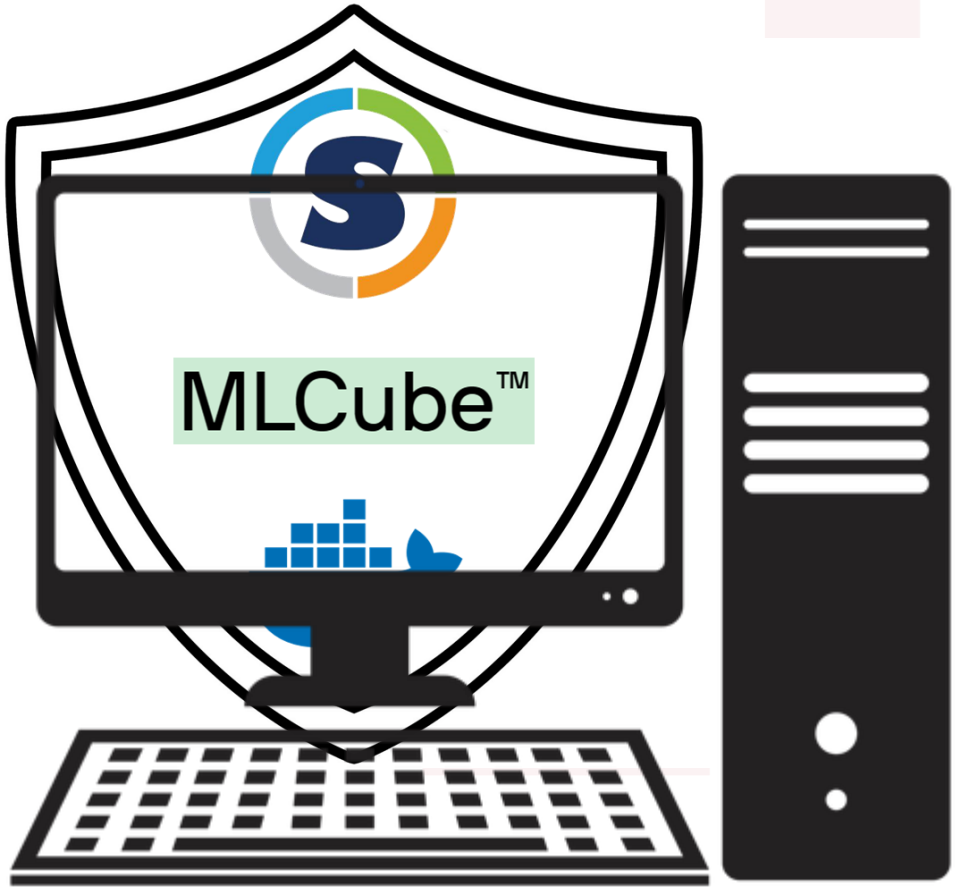
“Free” benefits of GANDLF

- Automatic Model Optimization [1]

- Easy translation to Federated Studies
  - Federated Learning using OpenFL [2]
  - Federated Evaluation using MedPerf [3]

- Support for secure containerization
  - MLCube
  - Singularity
  - Docker (with incorporated rail-guards)

- Automatic Model Deployment
  - Single-command use



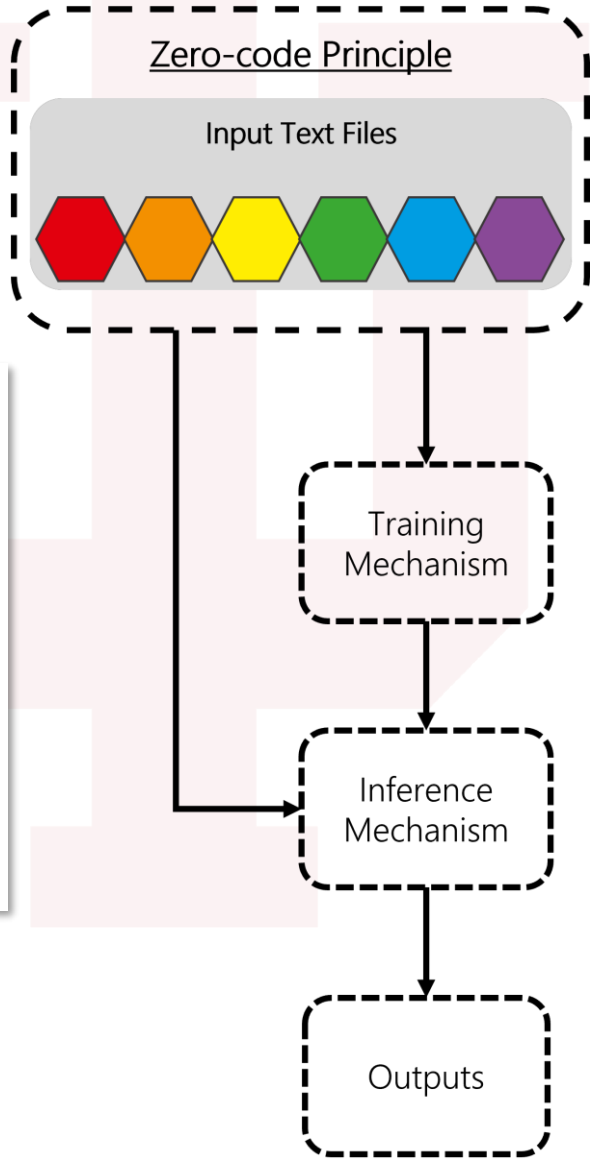
[1] Gorbachev et al.; DOI:10.1109/ICCVW.2019.00104
[2] Foley et al.; DOI:10.1088/1361-6560/ac97d9
[3] Karargyris et al.; DOI:10.48550/arXiv.2110.01406
[4] Cardoso et al.; DOI:10.48550/arXiv.2211.02701

Zero/Low Code Principle

- For novel experimental design
  - GaNDLF allows one to obtain quick baseline results
  - Head into publication cycle sooner

```
# Choose the model parameters here
model:
  {
    dimension: 3, # the dimension of the model and dataset: defines dimensionality of computations
    base_filters: 30, # 30 is for a GPU with 11GB VRAM - can be decreased/increased appropriately
    architecture: resunet, # options: unet, resunet, fcn, uinc
    final_layer: sigmoid, # can be either sigmoid, softmax or none (none == regression)
    class_list: '[1||2||4,1||4,4]', # Set the list of labels the model should train on and predict
    amp: True, # Set if you want to use Automatic Mixed Precision for your operations or not - options: True, False
    # n_channels: 3, # set the input channels - useful when reading RGB or images that have vectored pixel types
  }
```

- All of these result in 1+ **citations**!



Results and Future Goals

Results from Main Paper

Task	Organ	Application	Dims	Input Modalities (number):type	Output Classes	Architecture	Metric	
							Type	Average Value
Segmentation								
Regression								
Classification								

Community-driven effort

Siddhesh P. Thakur
Ibrahim Hamamci
Ujjwal Baid
Bhakti Baheti
Megh Bhalerao
Orhun Guley
Sofia Mouchtaris
David Lang
Spiridon Thermos
Karol Gotkowski
Camila Gonzalez
Caleb Grenko
Alexander Getka
Brandon Edwards

Micah Sheller
Junwen Wu
Deepthi Karkada
Ravi Panchumarthy
Vinayak Ahluwalia
Chunrui Zou
Vishnu Bashyam
Yuemeng Li
Babak Haghighi
Rhea Chitalia
Shahira Abousamra
Tahsin Kurc
Aimilia Gastounioti
Sezgin Er

Mark Bergman
Joel Saltz
Yong Fan
Prashant Shah
Anirban Mukhopadhyay
Sotirios A. Tsaftaris
Bjoern Menze
Christos Davatzikos
Despina Kontos
Alexandros Karargyris
Renato Umeton
Peter Mattson
Spyridon Bakas *

And multiple GitHub collaborators …

Open Source is the Way

Secure Containerization

Model Deployment in Low-resource Settings

Orchestration of AI Workflows

Federated Learning and Governance

MLCube™

OpenVINO™

MedPerf

OPENFL

GaNDLF

MONAI+

Torch/OpenCV

Numpy/Scikit

ITK/TiffSlide

SOTA Medical DL Algorithms

SOTA Computer Vision Algorithms

Stats, Math

I/O, Pre-process

Peer-reviewed Studies Facilitated by GANDLF

scientific reports

Improving privacy-preserving multi-faceted long short-term memory for accurate evaluation of encrypted time-series MRI images in heart disease

Lenka Čepová, Muniyandy Elangovan, Janjhyam Venkata Naga Ramesh, Mandeep Kaur Chohan, Amit Verma & Faruq Mohammad

Scientific Reports **14**, Article number: 20218 (2024) | Cite this article

Journals & Magazines > IEEE Internet of Things Journal > Early Access

communications engineering

Viewpoint | Open access | Published: 21 December 2023

Editors' Choice 2023

Miranda Vinay, Liwen Sang, Jianhua (Joshua) Tong, Or Perlman, Rosamund Daw, Carmine Galasso, Mengying Su, Damien Querlioz, Liangfei Tian, Anastasiia Vasylchenkova, Yu-Cheng Chen & Chaoran Huang

Communications Engineering **2**, Article number: 96 (2023) | Cite this article

A comprehensive review on federated

nature communications

nature > nature communications > articles > article

Article | Open Access | Published: 05 December 2022

Federated learning enables big data for rare cancer boundary detection

Sarthak Pati, Ujjwal Baid, Brandon Edwards, Micah Sheller, Shih-Han Wang, G. Anthony Reina, Patrick Foley, Alexey Gruzdev, Deepthi Karkada, Christos Davatzikos, Chiharu Sako, Satyam Ghodasara, Michel Bilello, Suyash Mohan, Philipp Vollmuth, Gianluca Brugnara, Chandrakanth J. Preetha, Felix Sahm, Klaus Maier-Hein, Maximilian Zenk, Martin Bendszus, Wolfgang Wick, Evan Calabrese, Jeffrey Rudie, ...

Spyridon Bakas + Show authors

Front Neurosci. 2024 May 20:18:1304191. doi: 10.3389/fnins.2024.1304191. eCollection 2024.

Prognostic stratification of glioblastoma patients by unsupervised clustering of morphology patterns on whole slide images furthering our disease understanding

Bhakti Baheti [1 2 3], Shubham Innani [1 2 3], MacLean Nasrallah [2 3], Spyridon Bakas [1 2 3 4 5]

Affiliations + expand

PMID: 38831756 PMCID: PMC11146603 DOI: 10.3389/fnins.2024.1304191

Future Development Goals

- **Address different data types**:
  - Genomics
  - EHR
  - Unstructured health data

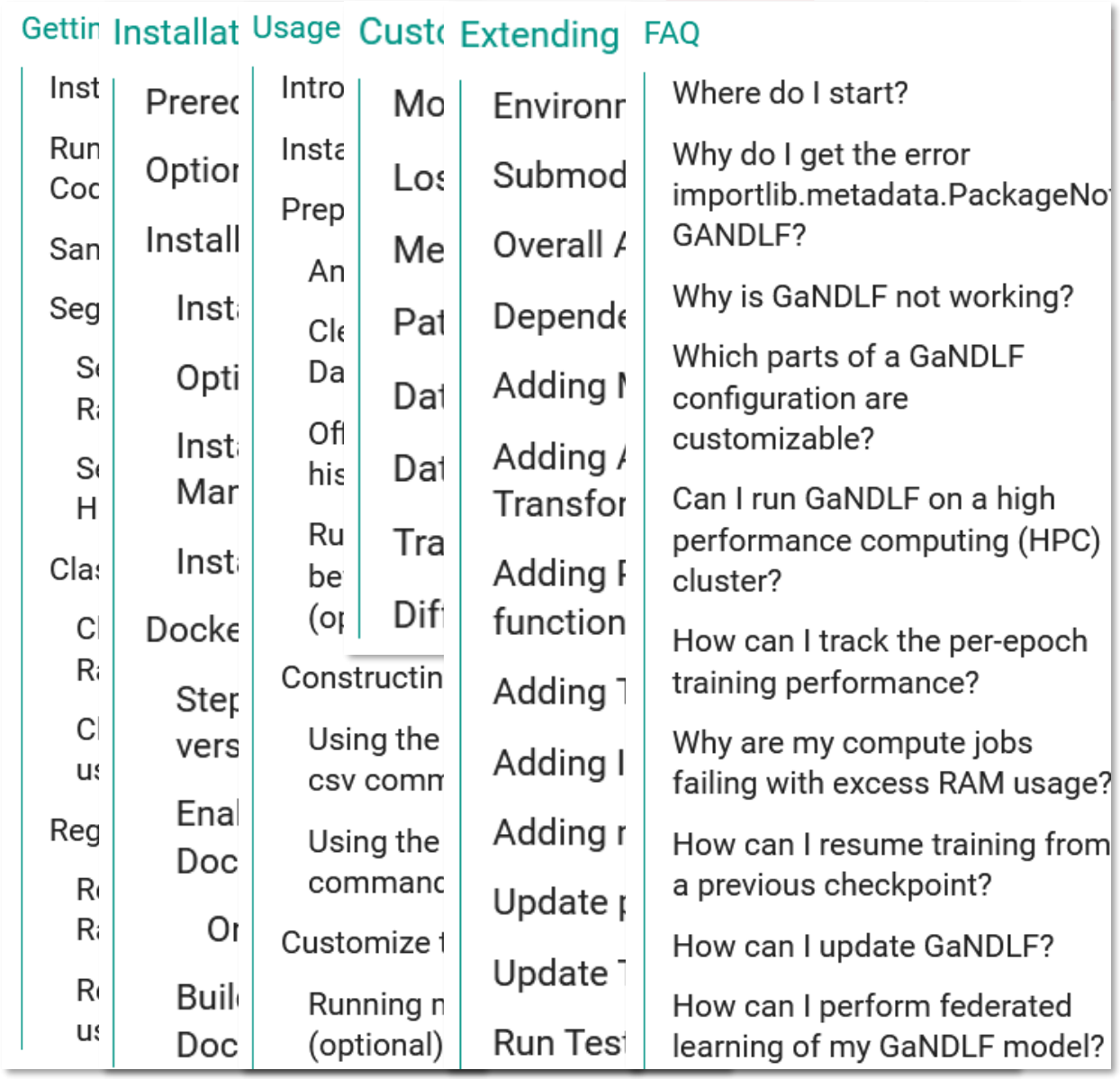
- **New workloads:**
  - Multi-instance learning
  - Semi-/Un-supervised learning
  - Multi-modal learning

Technical Specifics and Conclusions

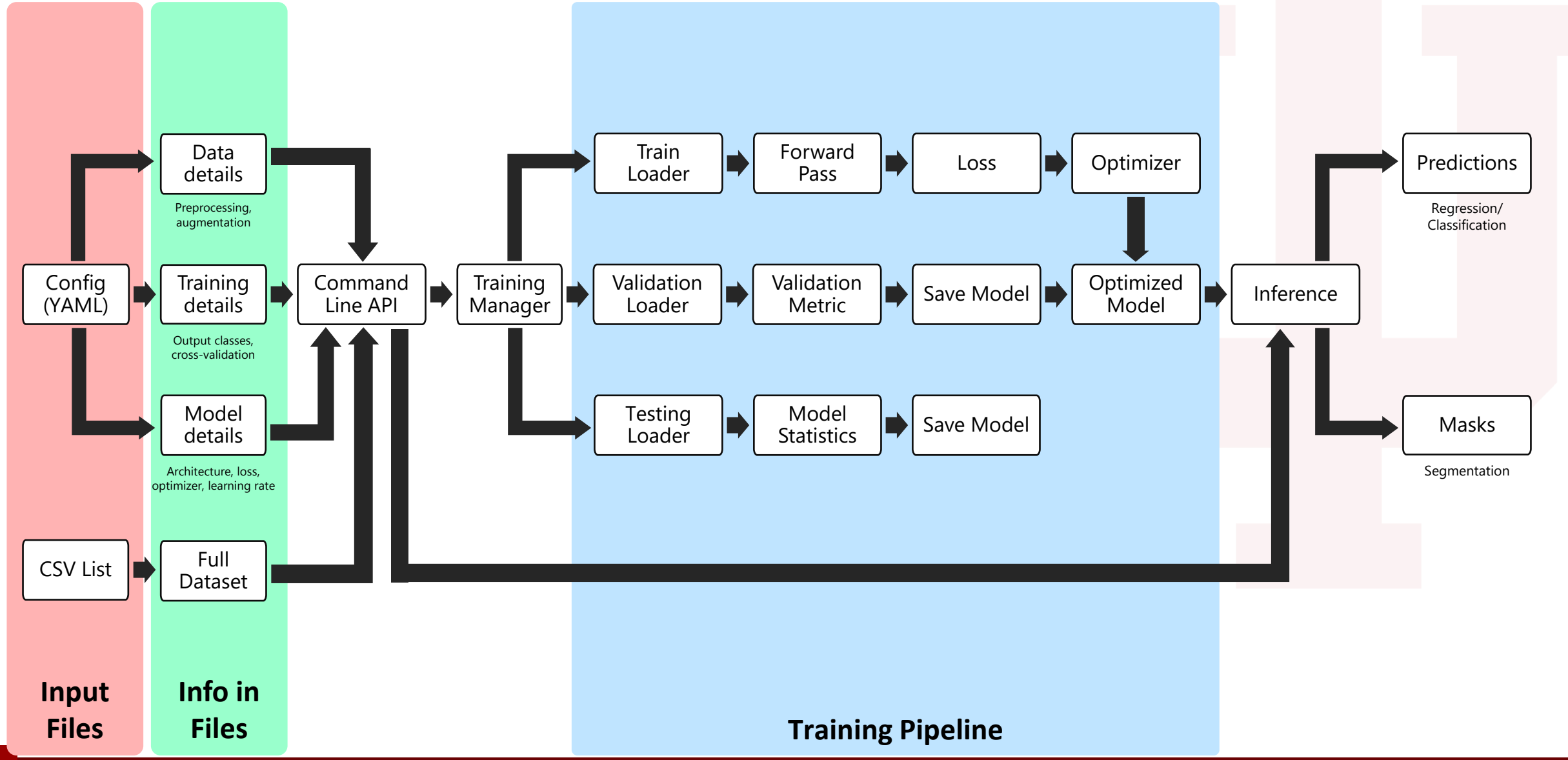
Documentation Links

- Main Starting point: gandlf.org

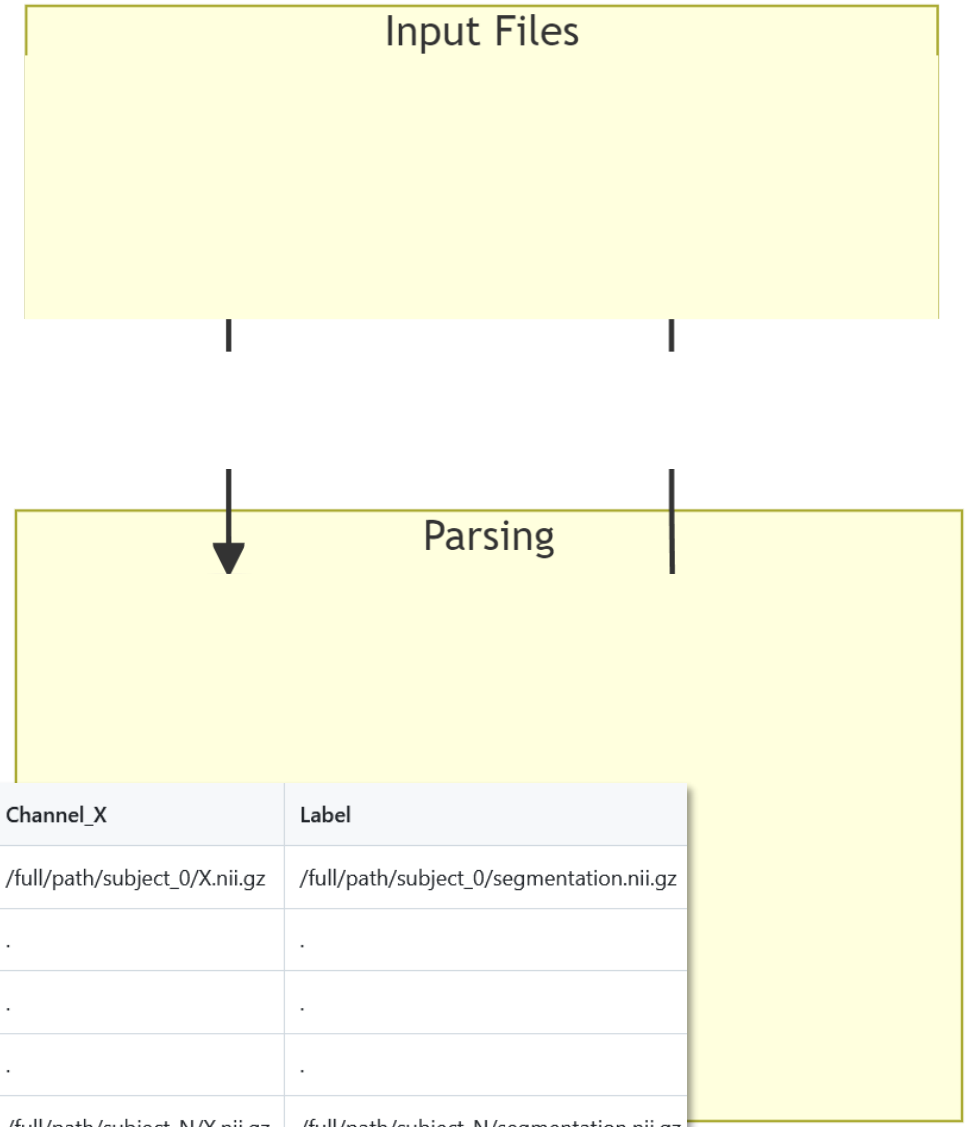
- Table of Contents for documentation
  - Getting Started
  - Application Setup
  - Usage
  - Customize the training and inference
  - Extending GaNDLF
  - FAQ



Overall Workflow



The I/O Mechanism

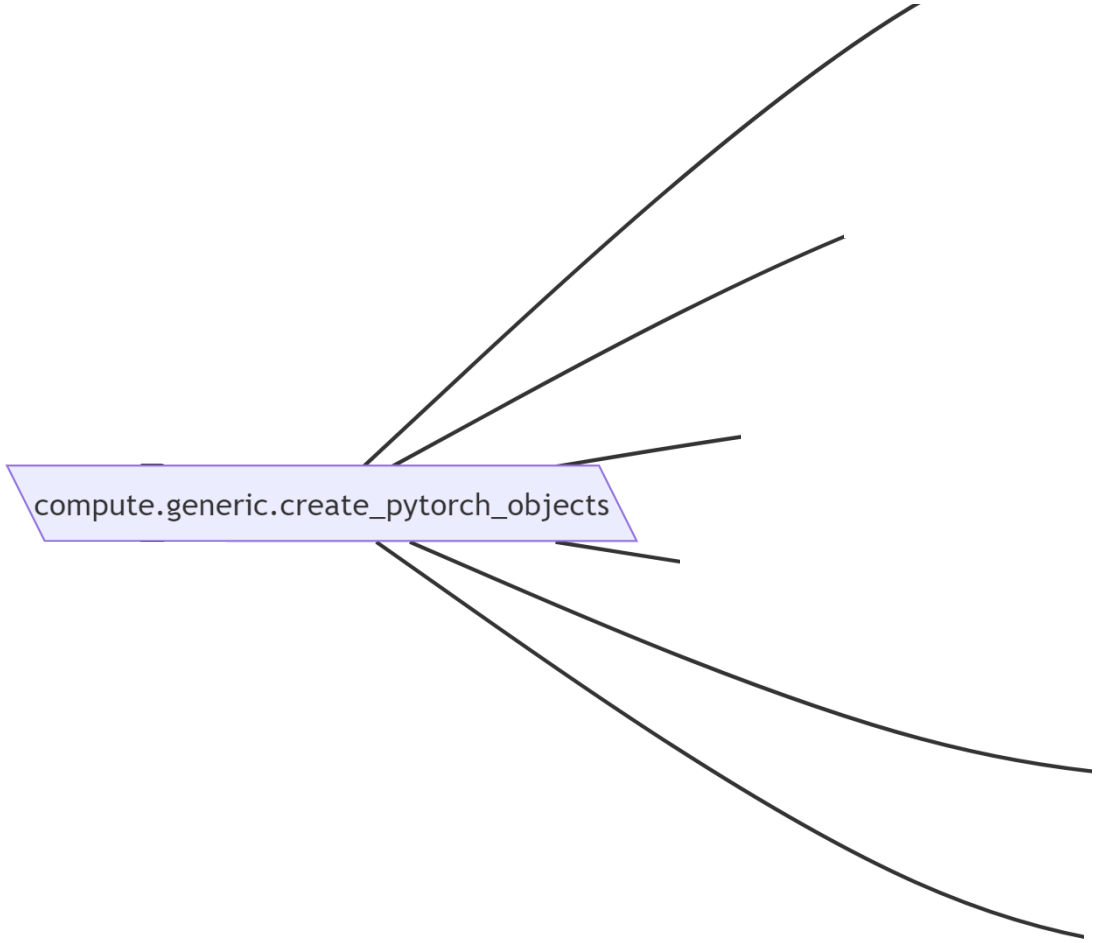



```
# Choose the model parameters here
model:
  {
    dimension: 2,
    architecture: vgg11,
    batch_norm: True,
    final_layer: None,
    amp: False,
    n_channels: 3,
    class_list: ["0", "1", "2", "3"],
  }
```

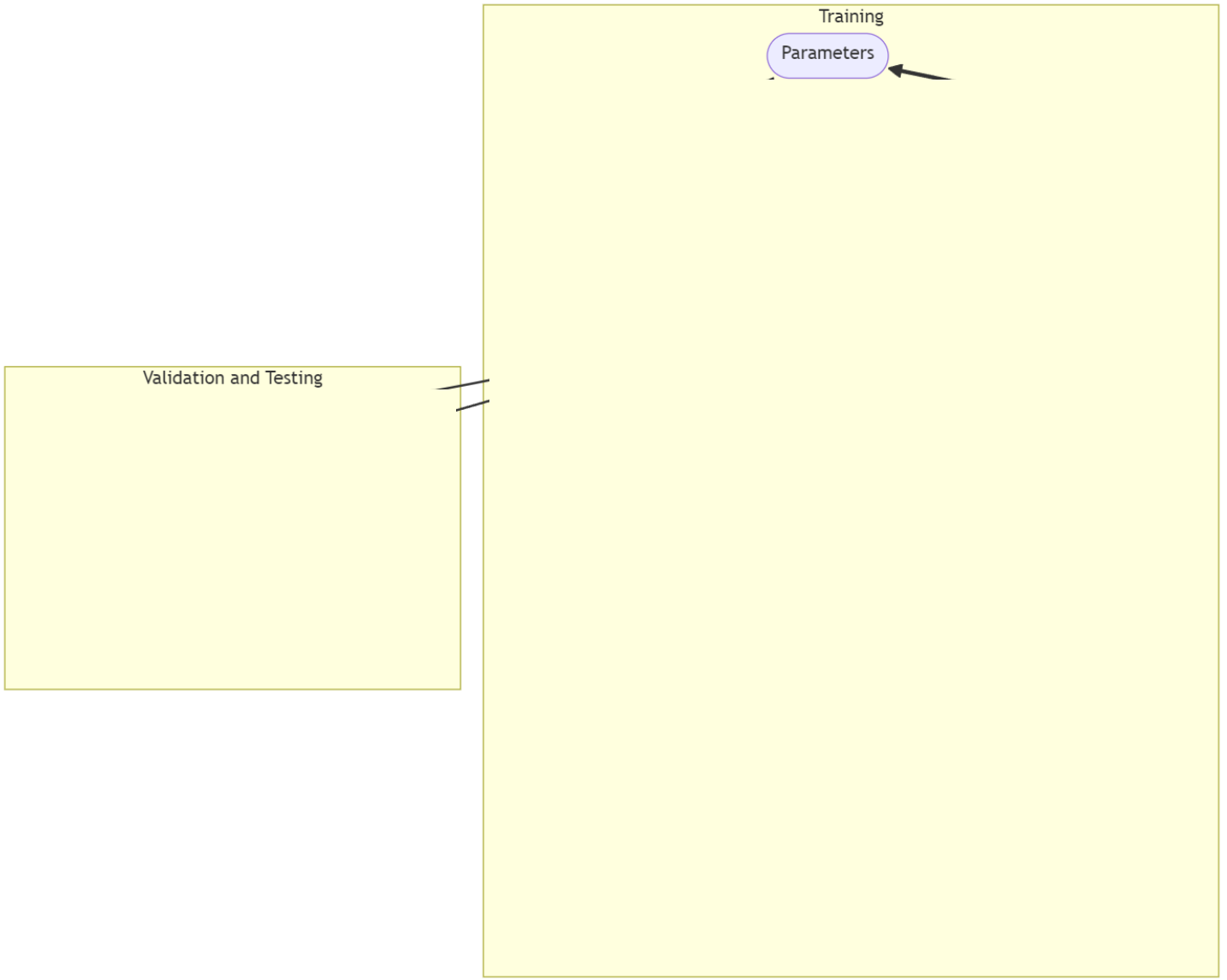
SubjectID	Channel_0	Channel_1	...	Channel_X	Label
subject_000	/full/path/subject_0/0.nii.gz	/full/path/subject_0/1.nii.gz	...	/full/path/subject_0/X.nii.gz	/full/path/subject_0/segmentation.nii.gz
.	.	.	...	.	.
.	.	.	...	.	.
.	.	.	...	.	.
subject_N	/full/path/subject_N/0.nii.gz	/full/path/subject_N/1.nii.gz	...	/full/path/subject_N/X.nii.gz	/full/path/subject_N/segmentation.nii.gz

Top Level Parsing

Main Function to Create Compute Objects



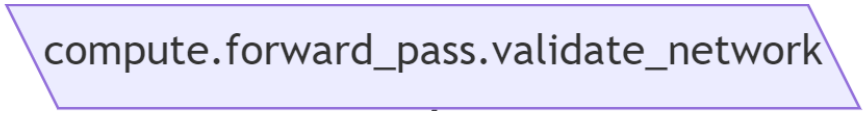
Training



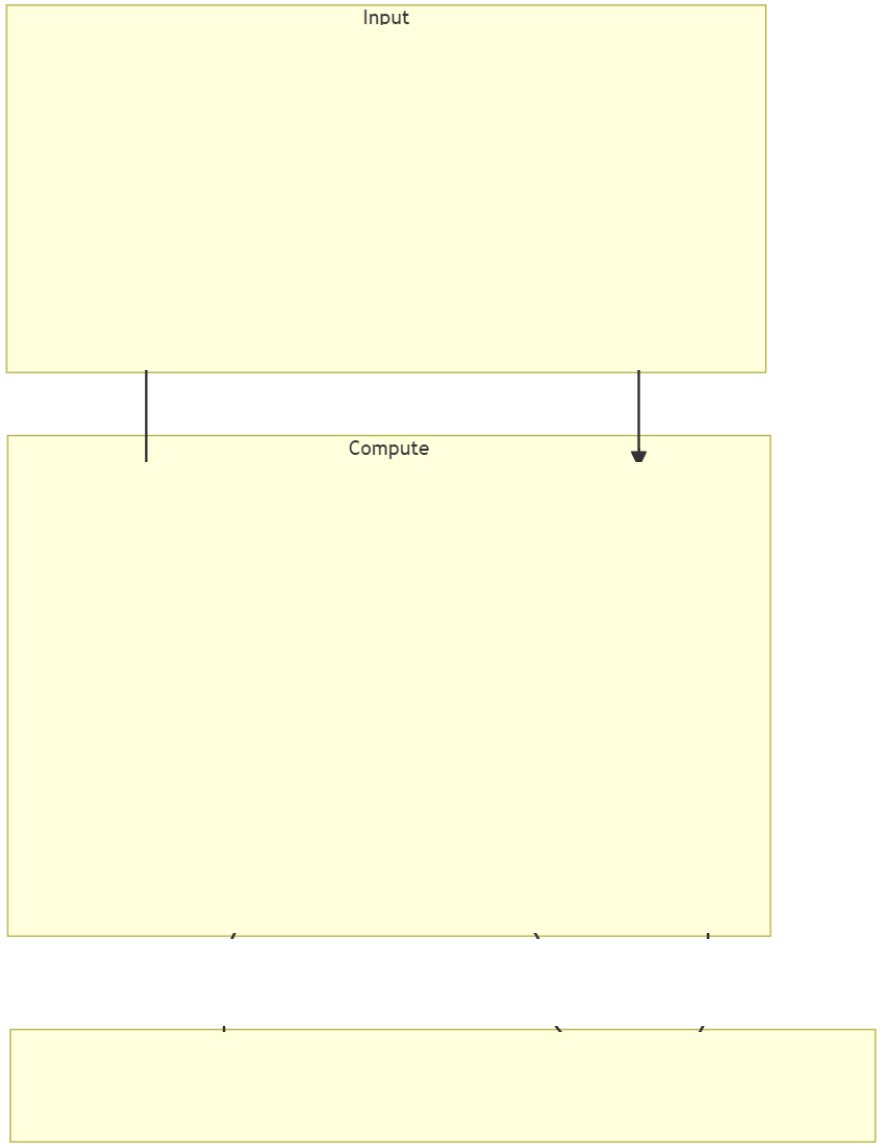
Validation

compute.forward_pass.validate_network

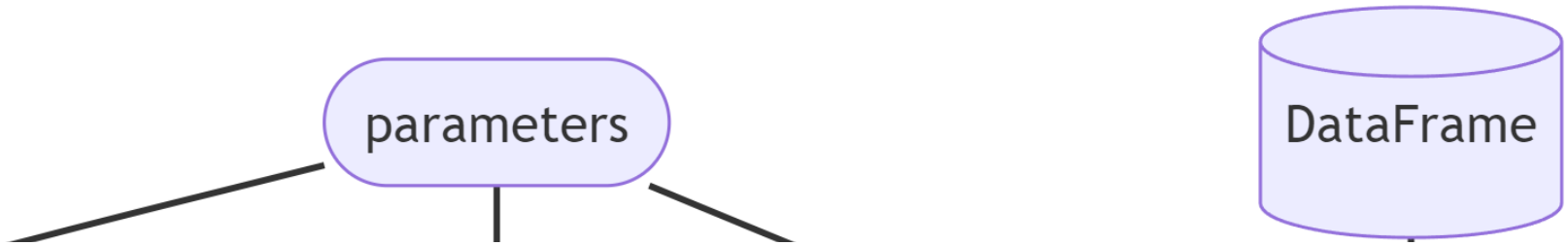
Inference



The `step` Routine



The `data` submodule



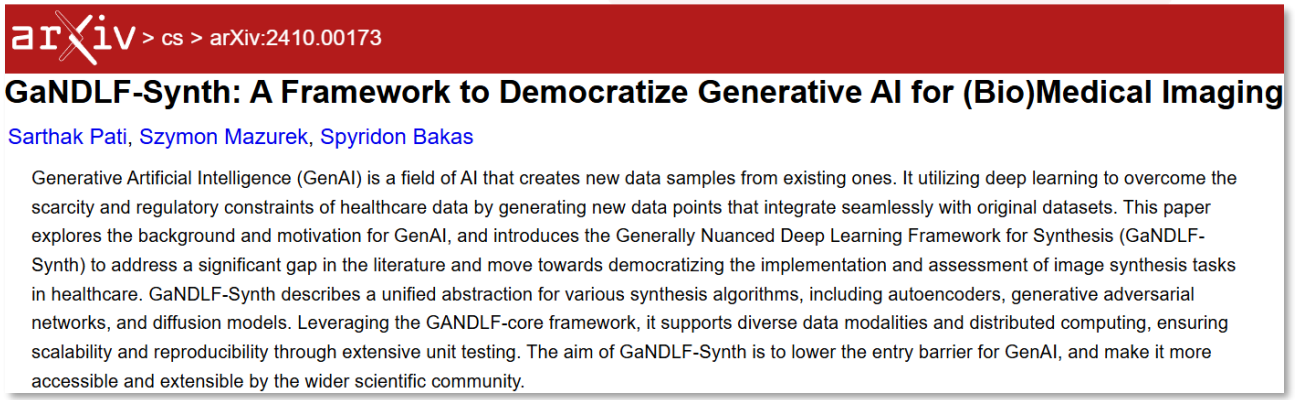
Announcements

Community Engagements

- We now have full integration with MONAI.
- We have full integration with Hugging Face Hub.

Noteworthy Features

- Privacy-enabled training.
- Full-scale synthesis as part of the ***GaNDLF-Synth*** extension!


arXiv > cs > arXiv:2410.00173

GaNDLF-Synth: A Framework to Democratize Generative AI for (Bio)Medical Imaging

Sarthak Pati, Szymon Mazurek, Spyridon Bakas

Generative Artificial Intelligence (GenAI) is a field of AI that creates new data samples from existing ones. It utilizing deep learning to overcome the scarcity and regulatory constraints of healthcare data by generating new data points that integrate seamlessly with original datasets. This paper explores the background and motivation for GenAI, and introduces the Generally Nuanced Deep Learning Framework for Synthesis (GaNDLF-Synth) to address a significant gap in the literature and move towards democratizing the implementation and assessment of image synthesis tasks in healthcare. GaNDLF-Synth describes a unified abstraction for various synthesis algorithms, including autoencoders, generative adversarial networks, and diffusion models. Leveraging the GANDLF-core framework, it supports diverse data modalities and distributed computing, ensuring scalability and reproducibility through extensive unit testing. The aim of GaNDLF-Synth is to lower the entry barrier for GenAI, and make it more accessible and extensible by the wider scientific community.

Key Takeaways