Publicly Available

***AstroCLIP: Cross-Modal Pre-Training for Astronomical Foundation Models***

**Data**

* DESI Legacy Survey
  + MetaData
    - 41M images
    - 152 x 152 → cropped to 96 x 96
    - 197,979 pairs
    - grayscale/rgb

***Attention-gating for improved radio galaxy classification***

<https://arxiv.org/pdf/2012.01248.pdf>

**Data**

* VLA FIRST Survey: <https://www.cv.nrao.edu/first/>
  + Requires FTP access in order get access to dataset
* MiraBest Dataset: <https://zenodo.org/record/4288837>
  + MetaData:
    - 1256 images (batches with 157 images each)
    - 150 x 150 pixels
    - Seems like 1 channel
* FR-Deep Dataset: <https://zenodo.org/record/4715983>
  + FR-DEEP dataset from the VLA FIRST Survey
  + MetaData
    - 1360 images
    - 150 x 150 pixels
    - 1 channel

**Model**

* AstroAttention Model: <https://github.com/mb010/AstroAttention>
  + Provides the AstroAttention Model

***CLASSIFYING RADIO GALAXIES WITH CONVOLUTIONAL NEURAL NETWORK***

<https://arxiv.org/pdf/1705.03413.pdf>

**Model**

* Radio Galaxy Classification: <https://github.com/ArunAniyan/RadioGalaxyClassification>
  + Model for classifying radio galaxies with CNN

**Data Manipulation Tools**

***CNN Architecture Comparison for Radio Galaxy Classification***

<https://arxiv.org/pdf/2102.03780.pdf>

**Data**

* Kaggle Galaxy Zoo: <https://www.kaggle.com/competitions/galaxy-zoo-the-galaxy-challenge/data>
  + Dataset from Kaggle competition
  + Classifying galaxies
  + Metadata
    - 61578 images (training set) + 79975 images (test set)
    - 424 x 424 pixels
    - 3 channels (RGB)
* Request from author

**Model**

* Model Benchmark: <https://github.com/BurgerBecker/rg-benchmarker>
  + Benchmark for using CNN for radio galaxy classification

***Convolutional Deep Denoising Autoencoders for Radio Astronomical Images***

<https://arxiv.org/pdf/2110.08618>

**Data Manipulation**

* killMS: <https://github.com/saopicc/killMS>
  + Implements direction dependent calibration algorithms

***Convolutional neural network analysis for x-ray diffraction data: strain profile retrieval in ion beam modified materials***

<https://iopscience.iop.org/article/10.1088/2632-2153/acab4c/pdf>

**Model**

* MLST Code: <https://github.com/aboulle/MLST-XRD>
  + Python source code for MLST

**Data Manipulation Tools**

***Deep learning based detection of cosmological diffuse radio sources***

<https://academic.oup.com/mnras/article-pdf/480/3/3749/25519514/sty2102.pdf>

**Data**

* Cosmodeep Dataset: <https://cosmosimfrazza.myfreesites.net/cosmodeep-training-datasets>
  + Sample dataset of sky models for training CNN Cosmodeep
  + Metadata
    - 61 images (.fits) data
    - 2000 x 2000 images
    - images seem to have 1 channel

***Deep Learning for Image Sequence Classification of Astronomical Events***

<https://arxiv.org/pdf/1807.03869.pdf>

**Data**

* Simulated datasets of labeled image sequences
  + <https://www.lsst.org/scientists/simulations/catsim>
  + Database of stars, galaxies, and solar system objects
  + Metadata
    - 4000 x 4072 pixels
    - Grayscale
    - Generated by specific inputs so many possible images

***Deep Learning improves identification of Radio Frequency Interference***

<https://arxiv.org/pdf/2005.08992>

**Data**

* HIDE: <https://hide.readthedocs.io/en/latest/usage.html>
  + Simulate one day of time-ordered-data from the bleien 7m radio telescope
* MeerKAT Simulation Package: <https://github.com/PetchMa/MeerKATgen>
  + Radio telescope simulation package focused on performance
* Available upon request

**Data Manipulation Tool**

* Ivy Workflow Engine: <https://github.com/cosmo-ethz/ivy>
  + Package for a workflow engine
* SEEK: <https://seek.readthedocs.io/en/latest/>
  + Processing pipeline for single dish radio telescopes
* KAT SIGProc: <https://github.com/ska-sa/katsdpsigproc>
  + Signal processing package for KAT data
* AOFlagger: <https://gitlab.com/aroffringa/aoflagger>
  + Tool for finding and removing radio-frequency interference RFI
* RFI-Mitigation: <https://github.com/vafaei-ar/RFI-mitigation>
  + DL based package for RFI mitigation

***Deep modeling of Quasar Variability***

<https://arxiv.org/pdf/2003.01241.pdf>

**Model**

* CatalinaQSO: <https://github.com/yutarotachibana/CatalinaQSO_AutoEncoder>
  + Uses Autoencoder for model

**Data Manipulation**

* Dust Reddening Tool: <https://irsa.ipac.caltech.edu/applications/DUST/>
  + NASA tool for dust reddening

***Deep Radio Interferometric Imaging with POLISH: DSA-2000 and weak lensing***

<https://arxiv.org/pdf/2111.03249.pdf>

**Model**

* POLISH model: <https://github.com/liamconnor/polish-pub/>
  + Code for the POLISH model

**Data Manipulation**

***Deeply Uncertain: Comparing Methods of Uncertainty Quantification in Deep Learning***

***Algorithms***

<https://arxiv.org/pdf/2004.10710.pdf>

**Model**

* DeeplyUncertain Model: <https://github.com/deepskies/DeeplyUncertain-Public>
  + Code repository for the paper

***Evaluation of Probabilistic Photometric Redshifts***

<https://www.researchgate.net/profile/Massimo->

**Data**

* SDSS DR9 data access (through METAphOR paper): <https://www.sdss3.org/dr9/data_access/>
  + Stores three types of data: image, spectra, and catalog
  + Sky survey
  + Unknown (a lot)
  + 2048 x 1489
  + Channels: 1 channel
* Requested by Buzzard Flock Group

**Model**

* ANNZ: <https://github.com/IftachSadeh/ANNZ>
  + Uses regression and classification techniques

**Data Manipulation Tools**

* BPZ-Bayesian Photometric Redshifts: <https://www.stsci.edu/~dcoe/BPZ/>
  + Had BPZ model and visualization
* EAZY: <https://github.com/gbrammer/eazy-photoz>
  + Code for creating new redshift pseudo-data when there is not enough data
* AIMALZ: <https://github.com/aimalz/qp/>
  + Home of QP which handles PDF using parametrization
* LSSTDESC: <https://github.com/aimalz/qp/>
  + Analysis and plotting

***Fanaroff-Riley classification of radio galaxies using group-equivariant convolutional neural networks***

<https://arxiv.org/pdf/2102.08252.pdf>

**Data**

* MiraBest Dataset: <https://zenodo.org/record/4288837>
  + MetaData:
    - 1256 images (batches with 157 images each)
    - 150 x 150 pixels
    - Seems like 1 channel

**Data Manipulation**

* E2 CNN Libary: <https://github.com/QUVA-Lab/e2cnn>
  + Pytorch extension for equivariant deep learning

***Fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks***

<https://cds.cern.ch/record/2746032/files/ATL-SOFT-PUB-2020-006.pdf?version=3>

**Data**

* GEANT4 Tool for Simulating Data: <https://geant4.web.cern.ch/>
  + Simulation toolkit

***Feature Selection Strategies for Classifying High Dimensional Astronomical Data Sets***

<https://arxiv.org/pdf/1310.1976.pdf>

**Data**

* Catalina Real-Time Transient Survey: <http://crts.caltech.edu/>
  + CRTS data provided in real time
  + Metadata
    - A lot
    - 5280 x 5320 pixels
    - Grayscale (1 channel)

***FETCH: A deep-learning based classifier for fast transient classification***

<https://arxiv.org/pdf/1902.06343.pdf>

**Data**

* FETCH Dataset: <http://astro.phys.wvu.edu/fetch/>
  + Dataset used by the FETCH paper
* Data from Zenodo: <https://zenodo.org/record/3905437>
  + Dataset detailed in the entirety of the paper

**Model**

* FETCH Model and Package: <https://github.com/devanshkv/fetch>
  + Code for the FETCH model

**Data Manipulation Tools**

* Heimdall: <https://sourceforge.net/projects/heimdall-astro/>
  + GPU-accelerated pipeline for data
* Pysigproc: <https://github.com/devanshkv/pysigproc>
  + Tool for processing sigproc format files

***GWSkyNet-Multi: A Machine Learning Multi-Class classifier for LIGO-VIRGO Public Alerts***

<https://arxiv.org/pdf/2111.04015.pdf>

**Model**

* GWSkyNet Model: <https://github.com/GWML/GWSkyNet>
  + Model for the GWSkyNet model
* Grad\_CAM: <https://github.com/ramprs/grad-cam/>
  + Provides code for the paper

**Data Manipulation**

* Skymap statistics: <https://github.com/reedessick/skymap_statistics>
  + Module and executables for quantifying and comparing skymaps
* LIGO skymap: <https://lscsoft.docs.ligo.org/ligo.skymap/>
  + Tools for reading, writing, generating, and visualizing gravitational-wave probability

***Improving the reliability of photometric redshift for machine learning***

<https://arxiv.org/pdf/2108.04784.pdf>

**Data**

* CasJobs: <https://skyserver.sdss.org/CasJobs/>
  + Online workbench for emulating and enhancing local free-form query access
  + Metadata
    - Currently unable to find

**Model**

* COSMOS\_SOM: <https://github.com/ShrRa/COSMOS_SOM>
  + SOM-based anomaly detection
* MINISOM: <https://github.com/JustGlowing/minisom>
  + Mini version of the SOM algorithm

***A Machine Learning-based Direction-of-origin Filter for the Identification of Radio Frequency Interference in the Search for Technosignatures***

<https://arxiv.org/pdf/2108.00559.pdf>

**Model**

* Source Code: <https://github.com/UCLA-SETI-Group/doom/releases/tag/v1.0.1>
  + Source code for the paper

**Data Manipulation Tools**

***Machine learning for the Zwicky Transient Facility***

<https://arxiv.org/pdf/1902.01936>

**Data**

* Kowalski Model Data (from Kowalski Model Link): <https://github.com/dmitryduev/kowalski/tree/master/data>
  + Has ZTF data used for the Kowalski Model

**Model**

* Kowalski Model: <https://github.com/dmitryduev/kowalski>
  + Machine Learning Model
  + Has library for different tools used

**Data Manipulation Tools**

* AmpelProject: <https://github.com/AmpelProject>
  + Tool for processing large amounts of heterogeneous data

***Morphological classification of radio galaxies: Capsule Networks versus Convolutional Neural Networks***

<https://arxiv.org/pdf/1905.03274.pdf>

**Data**

* LOFAR Surveys: <https://lofar-surveys.org/releases.html>
  + Data releases from the LOFAR surveys

**Model**

* CapsNet-Keras: <https://github.com/XifengGuo/CapsNet-Keras>
  + The Keras implementation of CapsNet
* Conv Caps: <https://github.com/vlukic973/RadioGalaxy_Conv_Caps>
  + Python code for classifying radio galaxies using ConvNet and CapsNet architectures

***ParSNIP: Generative Models of Transient Light Curves with Physics-Enabled Deep Learning***

<https://arxiv.org/pdf/2109.13999.pdf>

**Data**

* Pan-STARRS1 Data: <https://outerspace.stsci.edu/display/PANSTARRS/>
  + Dataset for Pan-STARRS
  + MetaData
    - 2885 light curves with host-galaxy redshifts; 557 have spectroscopically-confirmed types
    - 60 Orthogonal Transfer Arrays devices (OTA); each device has 8 x 8 array of “cells”; single OTA format is 4846 x 4868 pixel array; each device has 64 cells where each cell is 590 x 598 pixels
    - Seems grayscale (1 channel) when processed through a bandpass; full display uses 3 channels
* PLAsTiCC Dataset from Kaggle: <https://www.kaggle.com/code/michaelapers/the-plasticc-astronomy-starter-kit>
  + Dataset from Kaggle for astronomy purposes
  + Metadata
    - Tabular data or time series
    - First table: 12 features; Second Table: 6 features
    - Test set: ~3.5M objects; Training Data: 8000 sources, maybe ~3.5M objects

***Photometric redshift estimation via deep learning***

<https://arxiv.org/pdf/1706.02467.pdf>

**Data**

* SDSS Dataset: <http://www.skyserver.org/myskyserver/>
  + DR7 database
  + Metadata
    - 105,783 spectroscopically confirmed quasars (7th edition); 87,822 (9th edition)
    - Downloaded using HiPS in 28 x 28 square pixel format
    - Similar to DR9, assume grayscale (1 channel) after applying *ugriz* filters

***Photometric redshifts for Quasars in multi band Surveys***

<https://arxiv.org/pdf/1305.5641.pdf>

**Data**

* SDSS Dataset: <http://www.skyserver.org/myskyserver/>
  + DR7 database
  + Metadata
    - 105,783 spectroscopically confirmed quasars (7th edition); 87,822 (9th edition)
    - Downloaded using HiPS in 28 x 28 square pixel format
    - Similar to DR9, assume grayscale (1 channel) after applying *ugriz* filters
* Galex Dataset: <http://www.galex.caltech.edu/researcher/data.html>
  + Imaging dataset
  + Metadata
    - About 98182 objects
    - 3840 x 3840 pixels
    - Grayscale (1 channel)
* WISE Dataset: <https://irsa.ipac.caltech.edu/Missions/wise.html>
  + Wide-field infrared survey explorer
  + Metadata
    - 18240 images
    - Images 437x438 pixels
    - Grayscale (1 channel)
    - 4-band calibrated

***Pulsar Candidate Identification Using Semi-Supervised Generative Adversarial Networks***

<https://arxiv.org/pdf/2010.07457.pdf>

**Data**

* HTRS-Survey: <https://sites.google.com/site/htrupublications/htru-discoveries>
  + Has primarily medlat and hilate
  + Metadata
    - >100 pulsar discoveries
    - Tabular data
* Required by request from author

**Model**

* UBC AI: <https://github.com/zhuww/ubc_AI/blob/master/how_to_train_your_AI.py>
  + Neural network model
* GAN Model: <https://github.com/vishnubk/sgan.git>
  + GAN model used in the paper

**Data Manipulation Tools**

* PRESTO: <https://github.com/scottransom/presto>
  + Analysis tool that can analyze pulsar observations
* SIGPROC: <https://sigproc.sourceforge.net/>
  + Tool for standardizing initial analysis of pulsar type data

***Quantifying Uncertainty in Deep Learning Approaches to Radio Galaxy Classification***

<https://arxiv.org/pdf/2201.01203.pdf>

**Data**

* MiraBest Dataset: <https://zenodo.org/record/4288837>
  + MetaData:
    - 1256 images (batches with 157 images each)
    - 150 x 150 pixels
    - Seems like 1 channel

**Model**

* Radio Galaxies BBB: <https://github.com/devinamhn/RadioGalaxies-BBB>

**Data Manipulation Tools**

<https://github.com/devinamhn/RadioGalaxies-BBB>

<https://doi.org/10.5281/zenodo.4288837>

***Quantum Machine Learning for Radio Astronomy***

<https://arxiv.org/pdf/2112.02655>

**Data**

* HTRU2 Dataset: <https://archive.ics.uci.edu/dataset/372/htru2>
  + Pulsar candidates collected during the HTRU survey
  + Metadata
    - 17898 entries
    - Tabular data (csv)

**Model**

* Full code: <https://github.com/kordham/qaum>
  + Code for recreating results

***Radio Galaxy Zoo: ClaRAN – a deep learning classifier for radio morphologies***

<https://arxiv.org/pdf/1805.12008>

**Data**

* Kaggle Galaxy Zoo: <https://www.kaggle.com/competitions/galaxy-zoo-the-galaxy-challenge/data>
  + Dataset from Kaggle competition
  + Classifying galaxies
  + Metadata
    - 61578 images (training set) + 79975 images (test set)
    - 424 x 424 pixels

**Model**

* Claran Model: <https://github.com/chenwuperth/rgz_rcnn/>
  + Code for claran model

***Radio Galaxy Zoo: machine learning for radio source host galaxy cross-identification***

<https://academic.oup.com/mnras/article-pdf/478/4/5547/25204505/sty1308.pdf>

**Data**

* Github.io: <https://radiogalaxyzoo.github.io/atlas-xid/>
  + Claimed to have data, model, and everything
  + Could not find it

**Model**

* Crowdastro Model: <https://github.com/chengsoonong/crowdastro>
  + The model for the experiment

***RAPID: Early Classification of Explosive Transients using Deep Learning***

<https://arxiv.org/pdf/1904.00014>

**Data**

* PLAsTiCC Dataset from Kaggle: <https://www.kaggle.com/code/michaelapers/the-plasticc-astronomy-starter-kit>
  + Dataset from Kaggle for astronomy purposes
  + Metadata
    - Tabular data or time series
    - First table: 12 features; Second Table: 6 features
    - Test set: ~3.5M objects; Training Data: 8000 sources, maybe ~3.5M objects

**Model**

* Space Doctor: <https://zenodo.org/record/8289325>
  + Library of spacedoctor code and others
  + May have data

***Separation of pulsar signals from noise using supervised machine learning algorithms***

<https://arxiv.org/pdf/1704.04659>

**Data**

* SPINN Training Data: <http://astronomy.swin.edu.au/~vmorello/>
  + HTRU Medlat Dataset used in SPINN
  + Metadata
    - 1196 known pulsars; 89996 candidates
    - Seems to be tabular

**Model**

* Pulsar ML Model: <https://github.com/shiningsurya/pulsarml>
  + Has multilayer perceptron among many other models

***Weight Pruning and Uncertainty in Radio Galaxy Classification***

<https://arxiv.org/pdf/2111.11654>

**Data**

* MiraBest Dataset: <https://zenodo.org/record/4288837>
  + MetaData:
    - 1256 images (batches with 157 images each)
    - 150 x 150 pixels
    - Seems like 1 channel

**Model**

* RadioGalaxies-BBB: <https://github.com/devinamhn/RadioGalaxies-BBB>
  + Uses Bayesian Neural Network

**Extras**

*Reliable Photometric Redshifts Using ML*

<https://github.com/ShrRa/COSMOS_SOM>

*AI-assisted super-resolution cosmological simulations*

<https://github.com/eelregit/map2map>

*AI Feynman: a Physics-Inspired Method for Symbolic Regression*

<https://github.com/SJ001/AI-Feynman>

<https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge>

<https://github.com/BurgerBecker/rg-benchmarker>

*Avocado: Photometric Classification of Astronomical Transients with Gaussian Process Augmentation*

<https://github.com/kboone/avocado>

<https://github.com/lsst/throughputs>

*Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach*

<https://github.com/scienceguyrob/GHVFDT>

*SPINN: a straightforward machine learning solution to the pulsar candidate selection problem*

<http://astronomy.swin.edu.au/~vmorello/>

*Unveiling the rarest morphologies of the LOFAR Two-metre Sky Survey radio source population with self-organised maps*

github.com/lofar-astron/PyBDSF

<http://www.astron.nl/citt/pybdsf/algorithms.html#grouping>

astron.nl/citt/pybdsf/write\_catalog.html#write-catalog

astron.nl/citt/pybdsf/algorithms.html#gaussian-fitting

<https://github.com/HITS-AIN/PINK>

<http://yann.lecun.com/exdb/mnist/>

<https://lofar-surveys.org/public/LOFAR_HBA_T1_DR1_merge_ID_v1.2.comp.fits>

Zooniverse.org

*Rapid sorting of radio galaxy morphology using Haralick features*

<https://github.com/KushathaNtwaetsile/HaralickFeatures>

*Realistic Evaluation of Deep Semi-Supervised Learning Algorithms*

<https://github.com/brain-research/realistic-ssl-evaluation>

*Radio Galaxy zoo: using semi-supervised learning to leverage large unlabelled data-sets for radio galaxy classification under data-set shift*

<https://github.com/inigoval/fixmatch>

<https://doi.org/10.5281/zenodo.4288837>

*How Good is the Bayes Posterior in Deep Neural Networks Really?*

<https://github.com/google-research/google-research/tree/master/cold_posterior_bnn>

*AXS: A framework for fast astronomical data processing based on Apache Spark*

<http://github.com/mjuric/lsd>

<https://github.com/apache/spark/pull/21109>

*The ANTARES Astronomical Time-domain Event Broker*

<https://noao.gitlab.io/antares/client/>