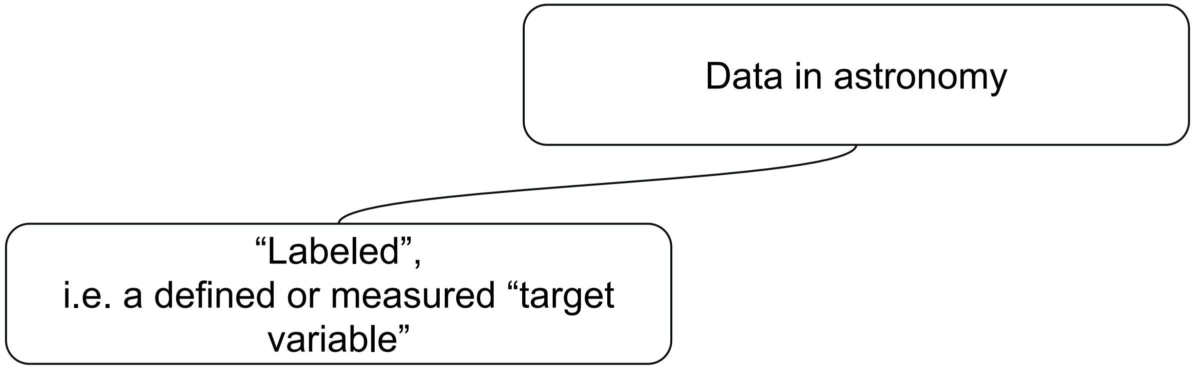

MACHING
12:29
DACHINE
LEARNING
02:9
MACHINE
LEARNING
IN ASTRONOW

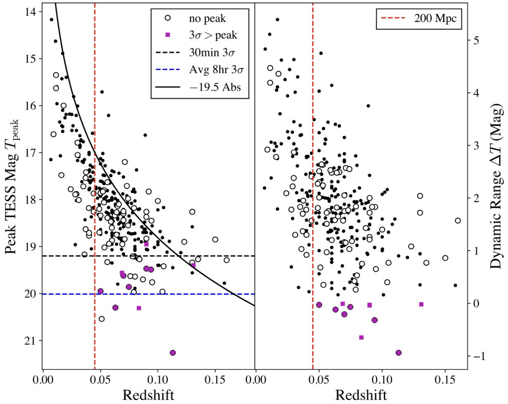
Data in Astronomy

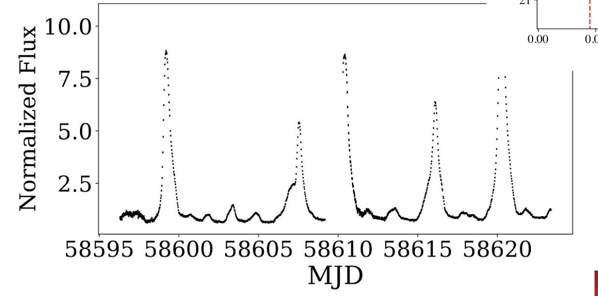
Data in astronomy


Data in astronomy
“Labeled”,
i.e. a defined or measured “target variable”

Data in astronomy

"Labeled",
i.e. a defined or measured "target variable"

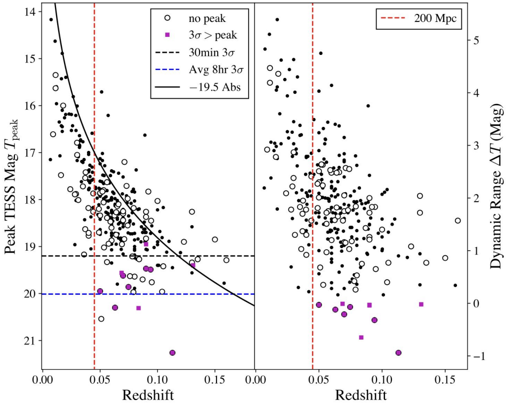


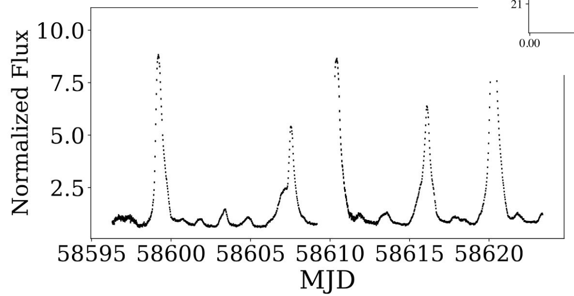


Data in astronomy

“Labeled”,
i.e. a defined or measured “target variable”

“Unlabeled”,
i.e. no defined “target variable”

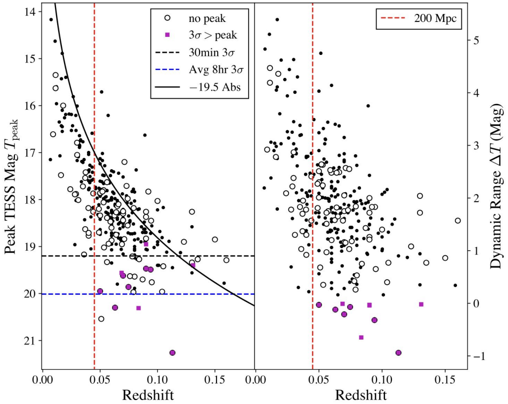


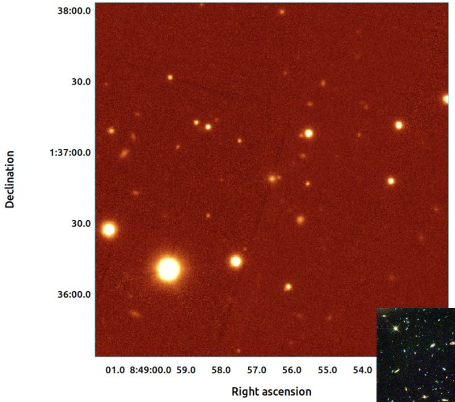


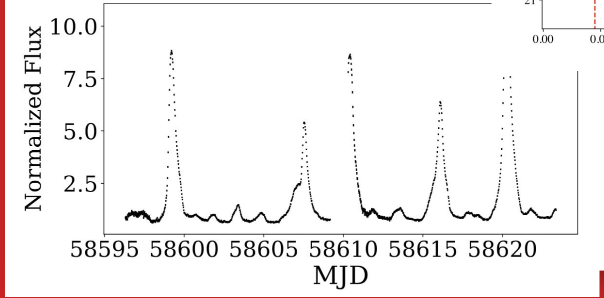
Data in astronomy

“Labeled”,
i.e. a defined or measured “target variable”

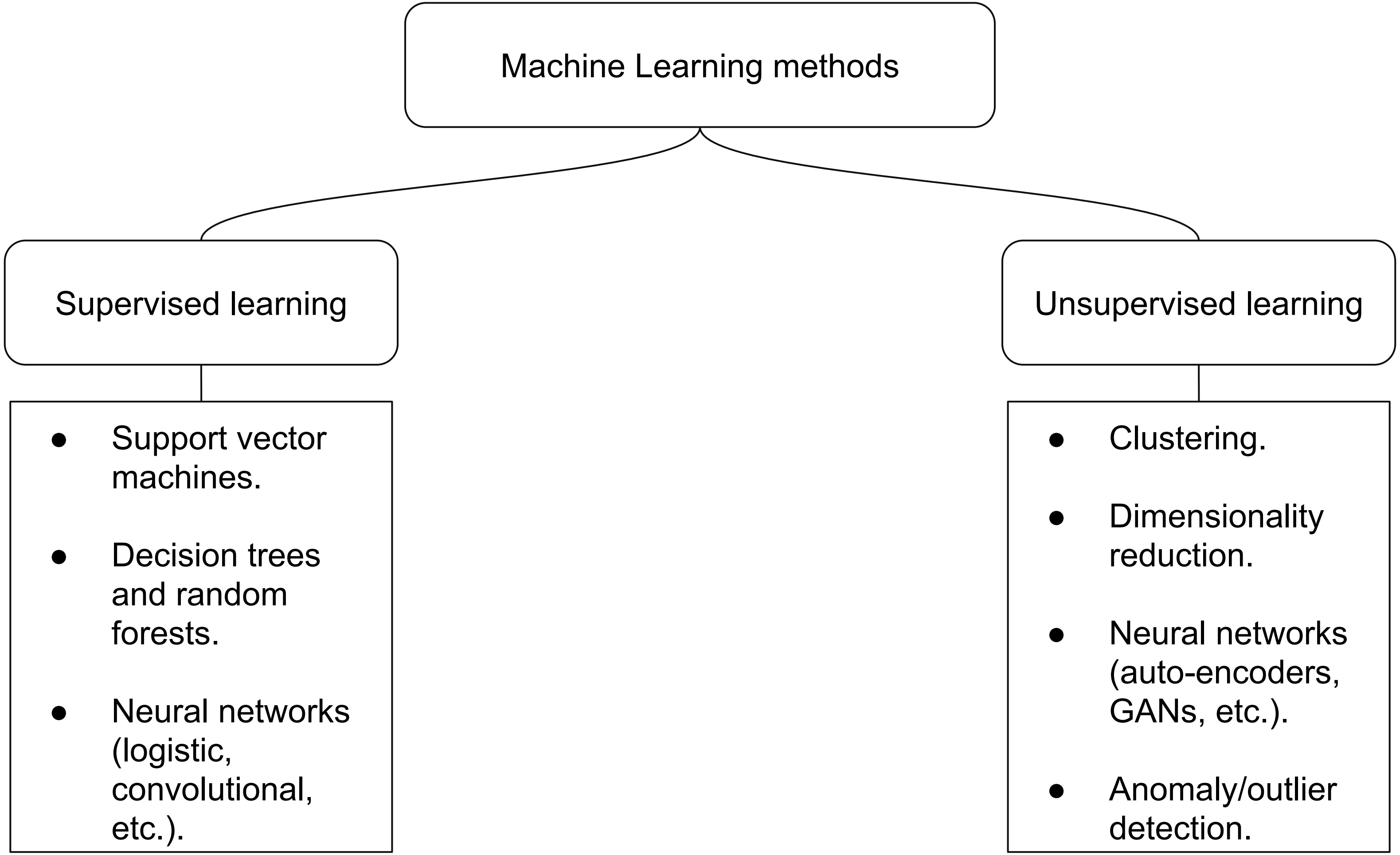
“Unlabeled”,
i.e. no defined “target variable”







Machine learning methods for astronomical data



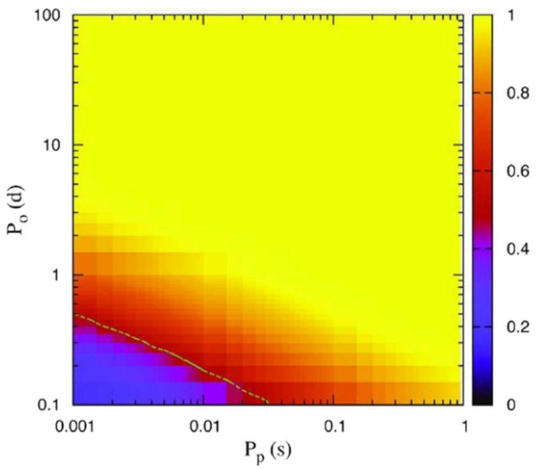
Machine Learning methods
Supervised learning
Unsupervised learning
● Support vector machines.
● Decision trees and random forests.
● Neural networks (logistic, convolutional, etc.).
● Clustering.
● Dimensionality reduction.
● Neural networks (auto-encoders, GANs, etc.).
● Anomaly/outlier detection.

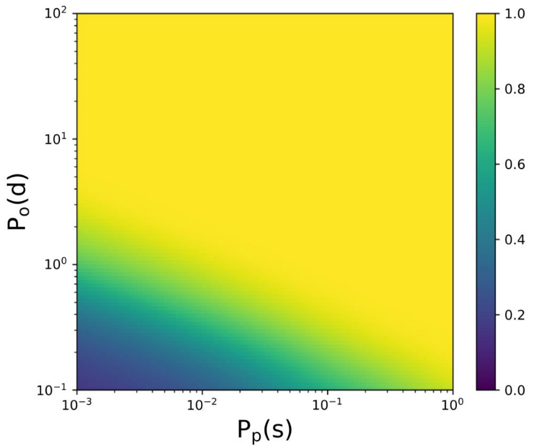
Supervised learning: Dense neural network regression

- Used a neural network to calculate the orbital degradation factor in the search for binary pulsars in radio data.

- Once trained, NN provides incredibly fast calculations, which are required in population synthesis studies.

$$\gamma_{2m}(\alpha_a, \alpha_v, T) = \frac{1}{T}\left|\int_0^T \exp\left[\frac{im\omega_p}{c}\left(r_l - r_{l0} - \alpha_a t^2 - \alpha_v t\right)\right] \mathrm{d}t\right|$$





Machine learning search for variable stars

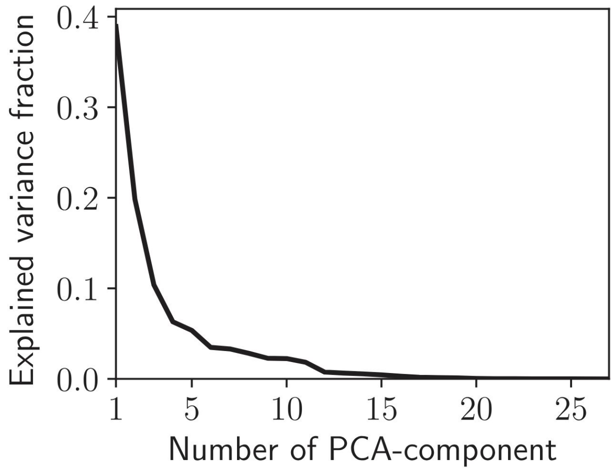
- Trained a broad range of ML algorithms on 168 objects (OGLE) using 18 variability indices.
- OGLE-II results: 205 candidates, of which 178 are real, and 13 are new discoveries.

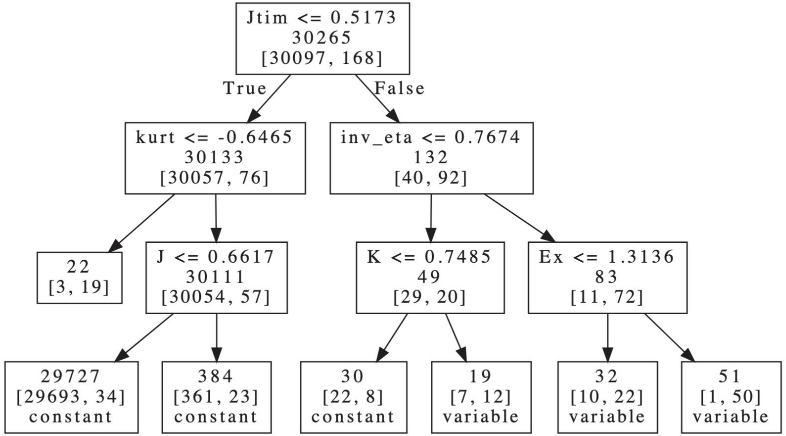
Type	LMC_SC19	LMC_SC20
Eclipsing binaries	36	54
Variable red giants (L/M/SR/ELL)	54	52
RR Lyrae-type variables	56	26
Cepheids (classical and Type II)	17	20
Blue irregular variables (GCAS/BE/QSO)	22	13
δ Scuti stars	1	3
Total	186	168

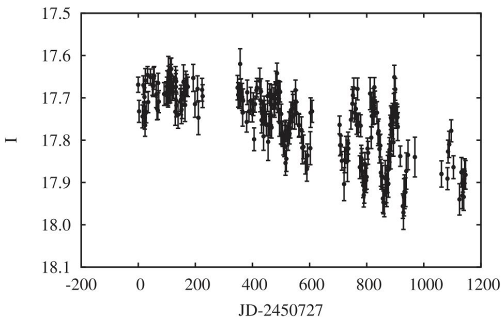
Index	Reference
Weighted standard deviation – σ	Kolesnikova et al. (2008)
Clipped σ – $\sigma_{\rm clip}$	Appendix A1
Median abs. deviation – MAD	Zhang et al. (2016)
Interquartile range – IQR	Sokolovsky et al. (2017)
Reduced χ^2 statistic – $\chi^2_{\rm red}$	de Diego (2010)
Robust median statistic – RoMS	Rose & Hintz (2007)
Norm. excess variance – $\sigma^2_{\rm NXS}$	Nandra et al. (1997)
Norm. peak-to-peak amp. – ν	Sokolovsky et al. (2009)
Autocorrelation – l_1	Kim et al. (2011)
Inv. von Neumann ratio – $1/\eta$	Shin et al. (2009)
Welch–Stetson index – $I_{\rm WS}$	Welch & Stetson (1993)
Flux-independent index – $I_{\rm fi}$	Ferreira Lopes et al. (2015)
Stetson's J index	Stetson (1996)
Time-weighted Stetson's $J_{\rm time}$	Fruth et al. (2012)
Clipped Stetson's $J_{\rm clip}$	Appendix A2
Stetson's L index	Stetson (1996)
Time-weighted Stetson's $L_{\rm time}$	Fruth et al. (2012)
Clipped Stetson's $L_{\rm clip}$	Appendix A2
Consec. same-sign dev. – CSSD	Shin et al. (2009)
S_B statistic	Figuera Jaimes et al. (2013)
Excursions – E_x	Parks et al. (2014)
Excess Abbe value – $\mathcal{E}_{\mathcal{A}}$	Mowlavi (2014)
Stetson's K index	Stetson (1996)
Kurtosis	Friedrich, Koenig & Wicenec (1997)
Skewness	Friedrich et al. (1997)

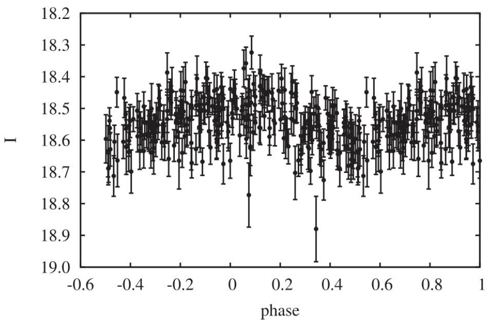
[Pashenko et al., 2018.]

Machine learning search for variable stars





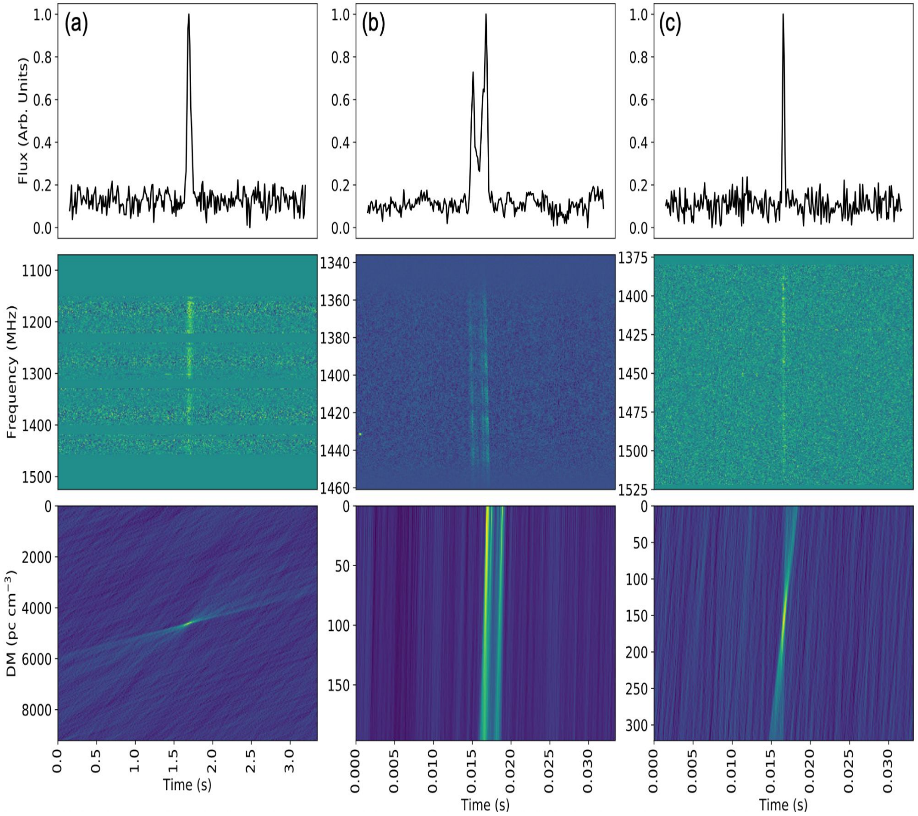




Agarwal et al., 2020

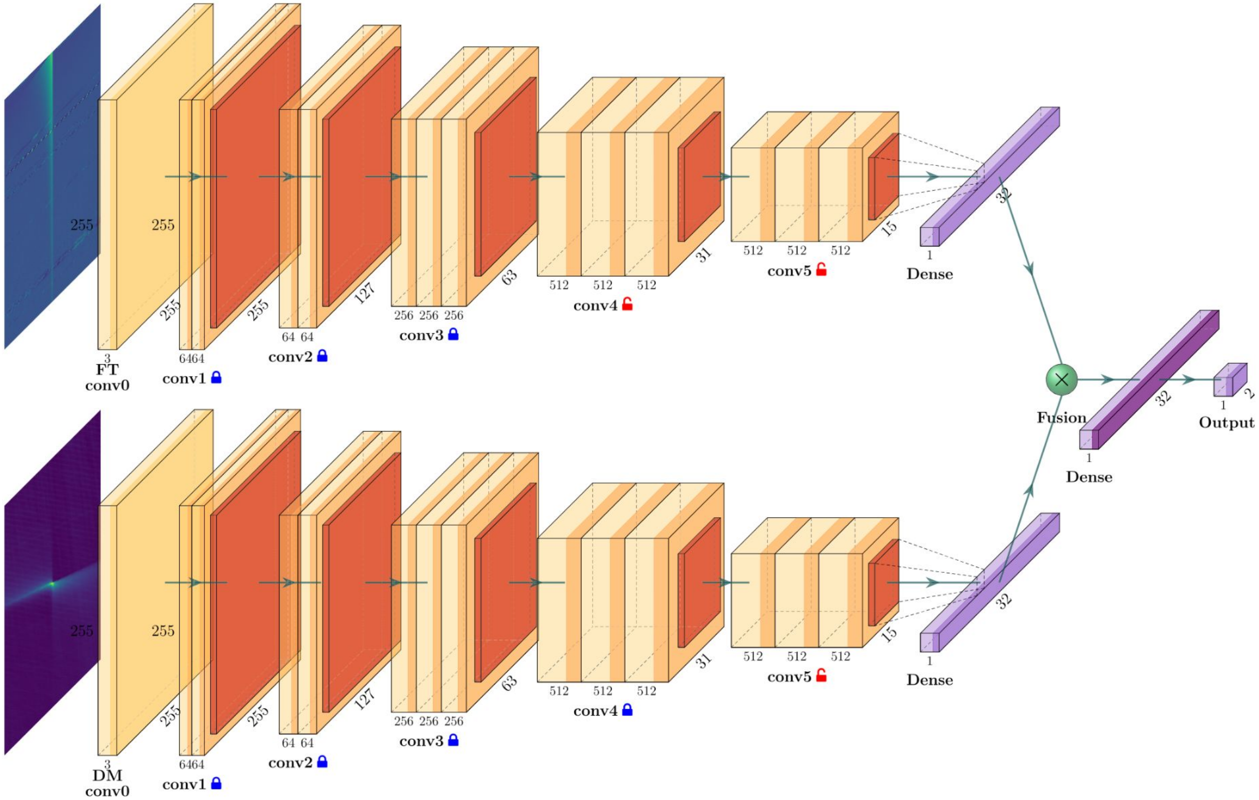
Transfer learning for classification

- Train a deep convolutional neural network to detect fast radio bursts, a class of radio transients.
- Bulk of neural network design leveraged image classifiers pre-trained on ImageNet like *ResNet50, VGG16, DenseNet.*



Agarwal et al., 2020

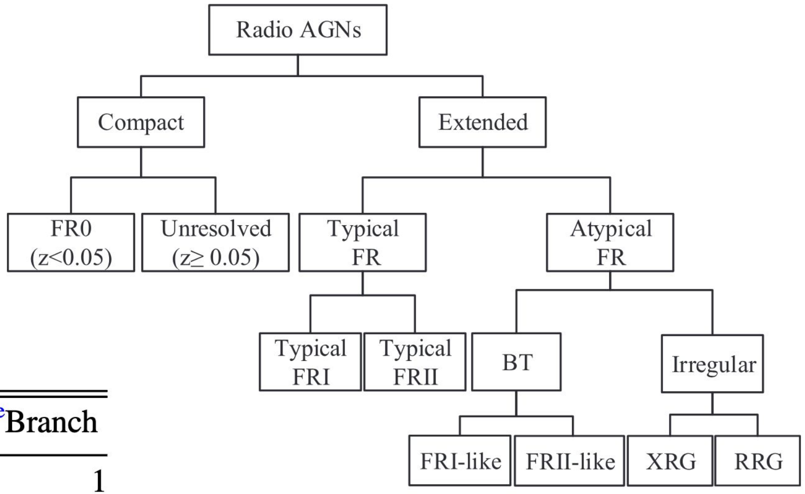
Transfer learning with classification



[Ma et al., 2019]

Classification with autoencoders

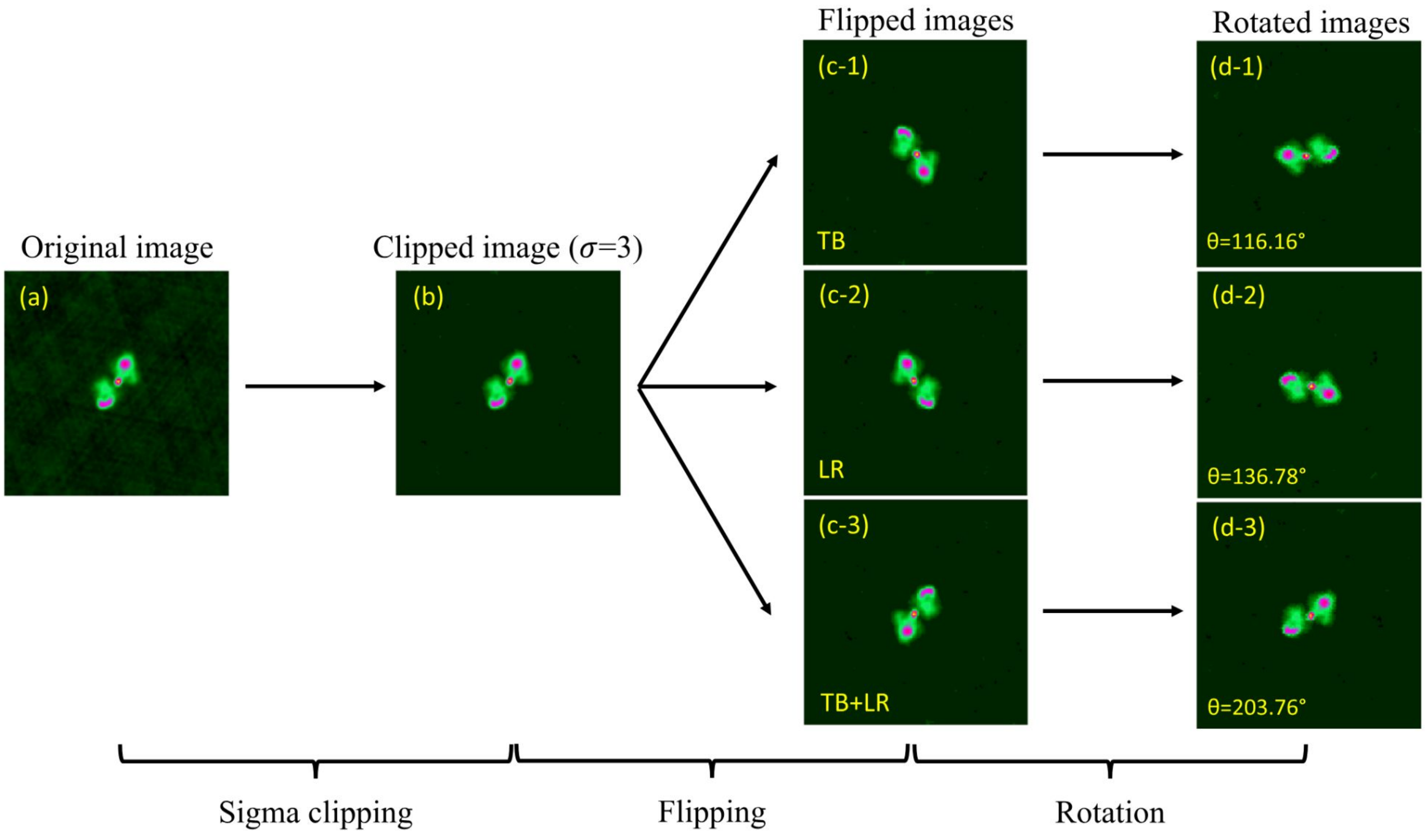
- Design a convolutional auto-encoder to classify radio AGN into six types: FR I/II, FR I/II-like bent-tailed, X-shaped, ring-like
- Clever use of augmentation to increase their training set.



AGN Type	Label	[a]$N_{\mathrm{Trn+Val}}$	[b]N_{Tst}	[c]R_{Aug}	[d]N_{Aug}	[e]Branch
compact	1	302	75	64	19,328	1
FRI	2	169	42	29	4901	1, 2, 3
FRII	3	345	86	14	4830	1, 2, 3
BT	4	245	62	20	4900	1, 2, 4, 5
XRG	5	67	17	37	2479	1, 2, 4, 6
RRG	6	26	6	94	2444	1, 2, 4, 6
Total	⋯	1154	288	⋯	38,882	⋯

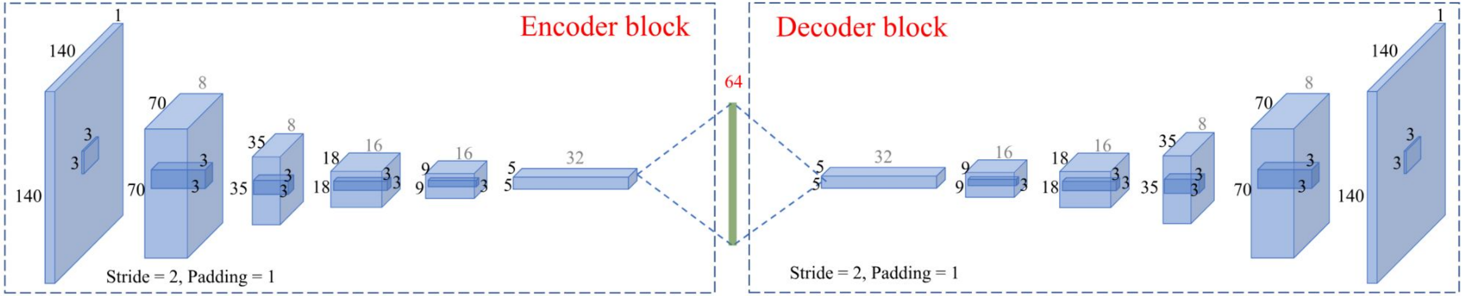
Classification with autoencoders

Ma et al., 2019

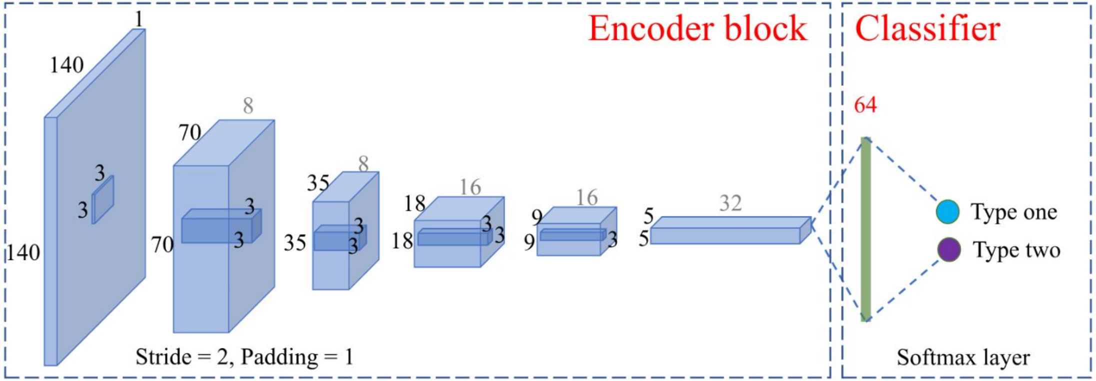


Classification with autoencoders

[Ma et al., 2019]



(a)



More interesting applications of ML to astronomy:

- More applications:
  - Bayesian neural networks
  - Distribution learning with normalizing flows.
  - Time series analysis with neural networks.
  - LLMs in astronomy? AstroLlama, Slack chatbots grounded in arXiv, survey of ChatGPT and LLM use by astronomers. (also: next week's colloquium speaker)
  - Outlier detection

- Many algorithms already available in Python:
  - Scikit-learn
  - Tensorflow
  - Keras for deep learning.