

# Machine learning for mineral exploration: A data odyssey

Dr Rohitash Chandra

School of Mathematics and Statistics, UNSW Sydney

Dr Ehsan Farahbakhsh

EarthByte Group, School of Geosciences, University of Sydney

CSIRO Minerals, Kensington, Perth

9th September, 2022



**UNSW**  
SYDNEY

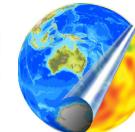


**DARE**

ARC Training Centre in  
Data Analytics for Resources & Environments



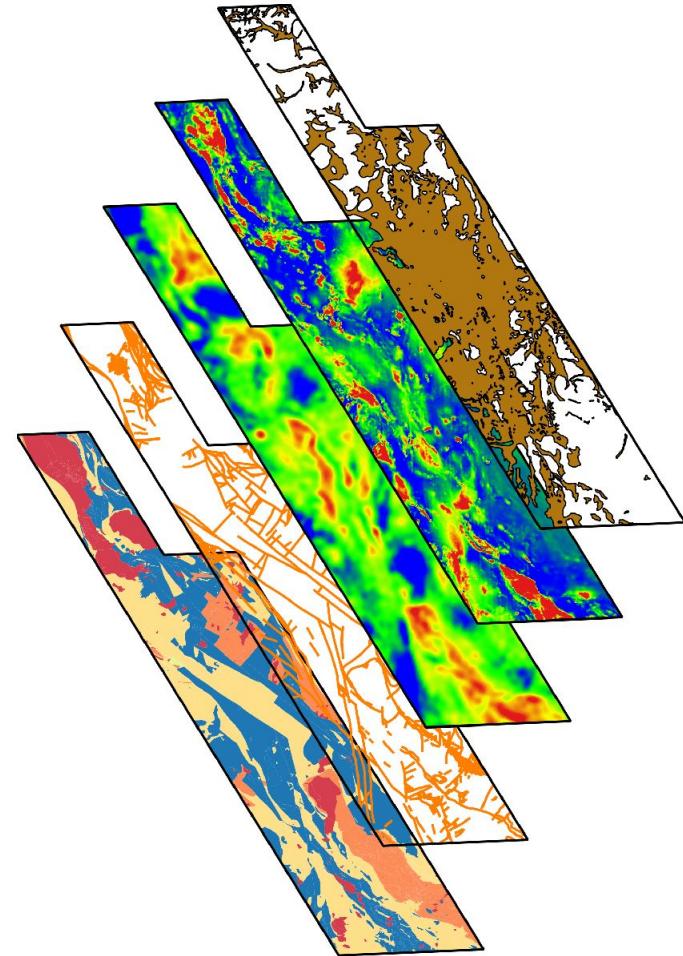
THE UNIVERSITY OF  
**SYDNEY**



**EarthBYTE**  
Building a Virtual Earth

# Overview

- Machine Learning-based framework for prospectivity mapping of critical minerals
- Autoencoders and Remote Sensing for mineral exploration
- Coupling deep learning with Plate Tectonics models for mineral exploration
- Convolutional Neural Networks for Lithological Mapping via Remote Sensing



# Machine Learning-based framework for prospectivity mapping of critical minerals

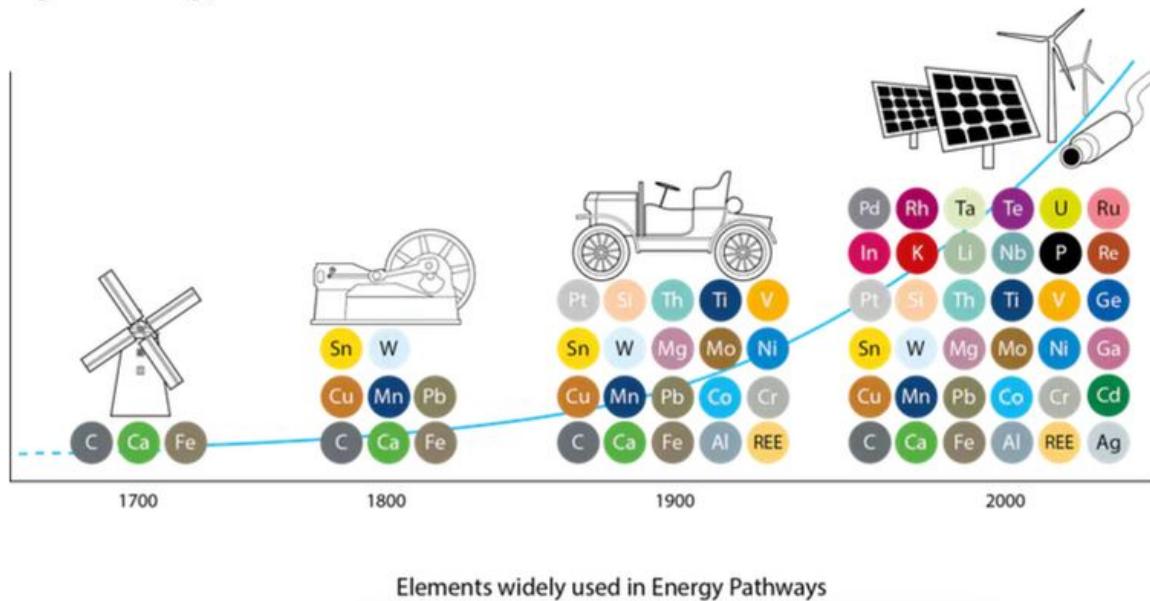
Ehsan Farahbakhsh



R. Dietmar Müller



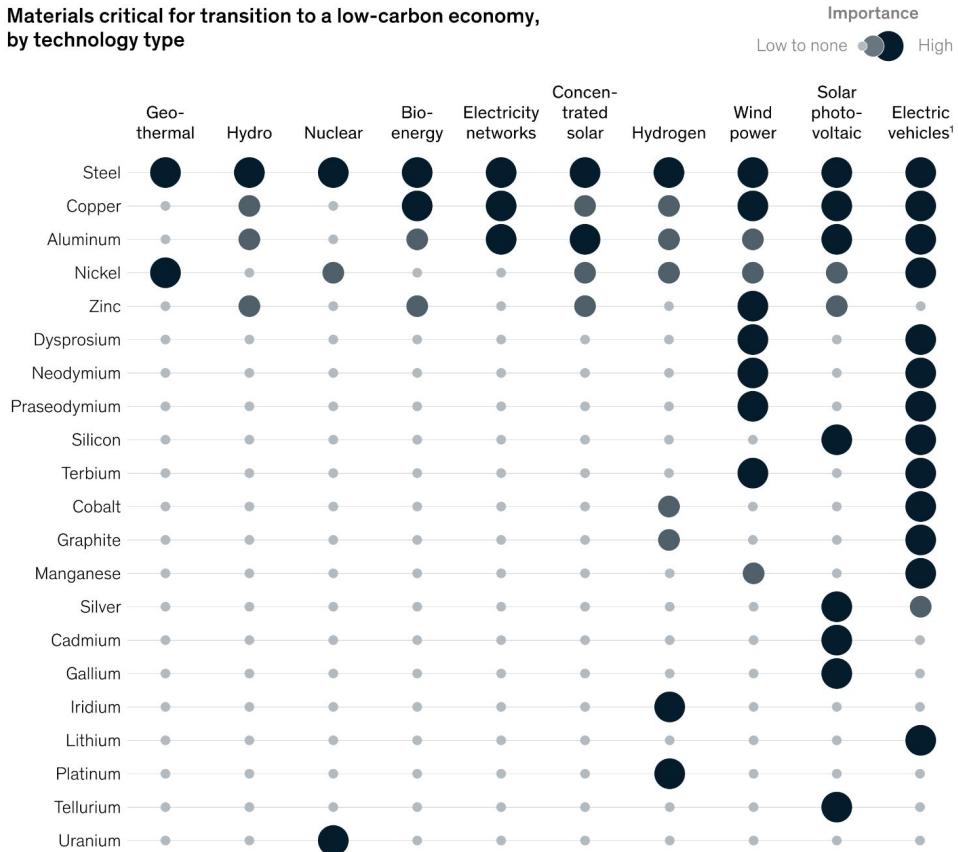
Ages of Energy



# Critical Minerals

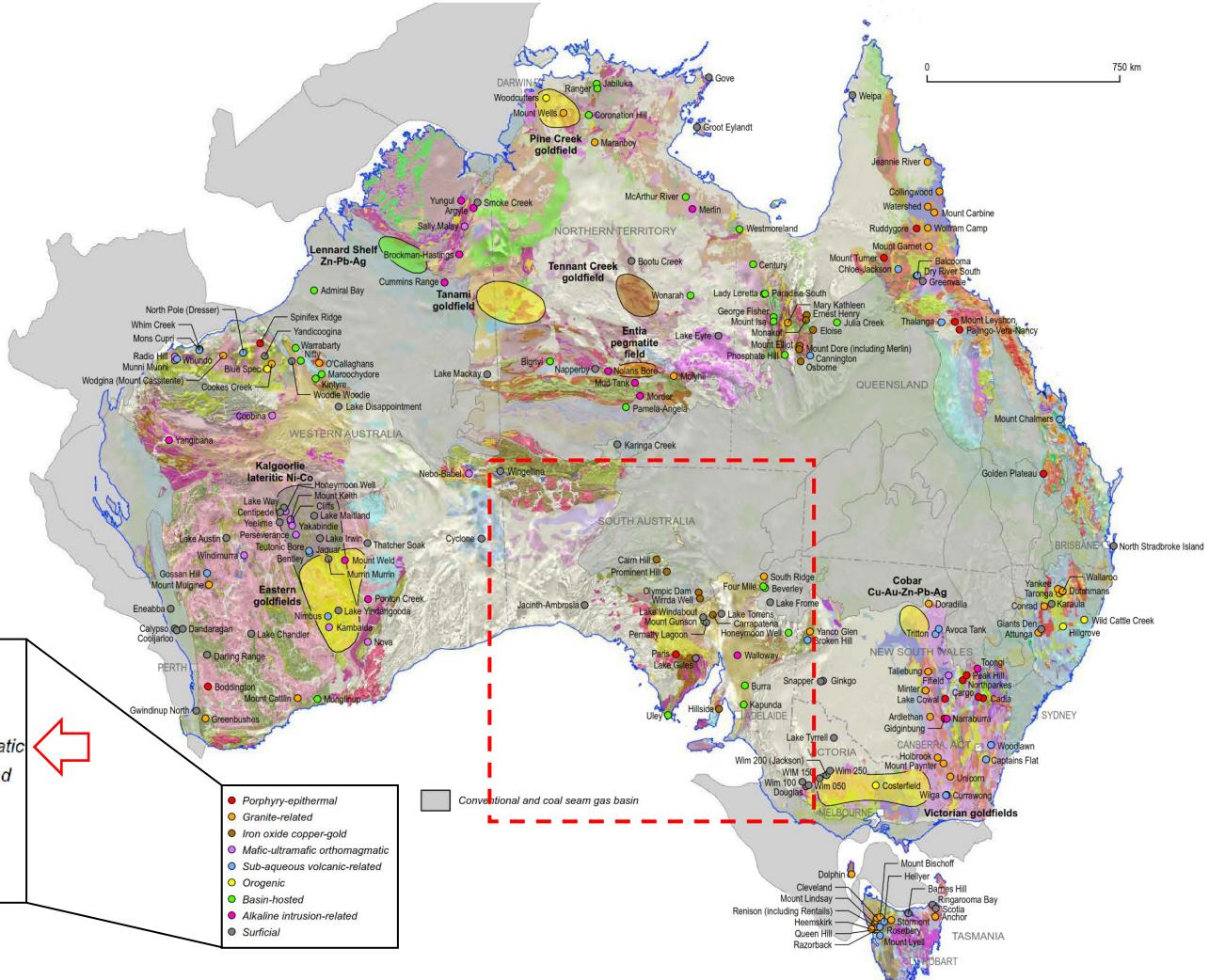
- Critical minerals are metallic or non-metallic minerals that are essential for the functioning of modern technologies, economies, or national security, and there is a risk that their supply chains could be disrupted.
- Many are used in low-emission technologies such as electric vehicles, wind turbines, solar panels, and rechargeable batteries.
- The Australian Government considers 26 resource commodities to be critical. These have been selected by assessing Australia's geological endowment and potential with global technology needs, particularly those of partner countries.

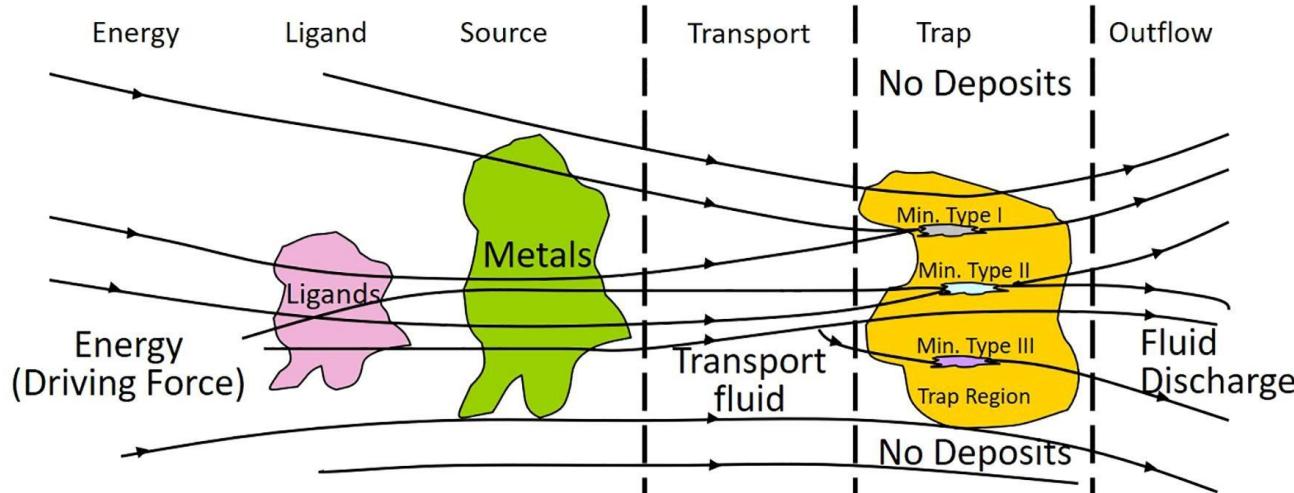
Materials critical for transition to a low-carbon economy,  
by technology type



<sup>1</sup>Includes energy storage.

Source: Critical raw materials for strategic technologies and sectors in the EU, A foresight study, European Commission, Mar 9, 2020; The role of critical minerals in clean energy transitions, IEA, May 2021; McKinsey analysis

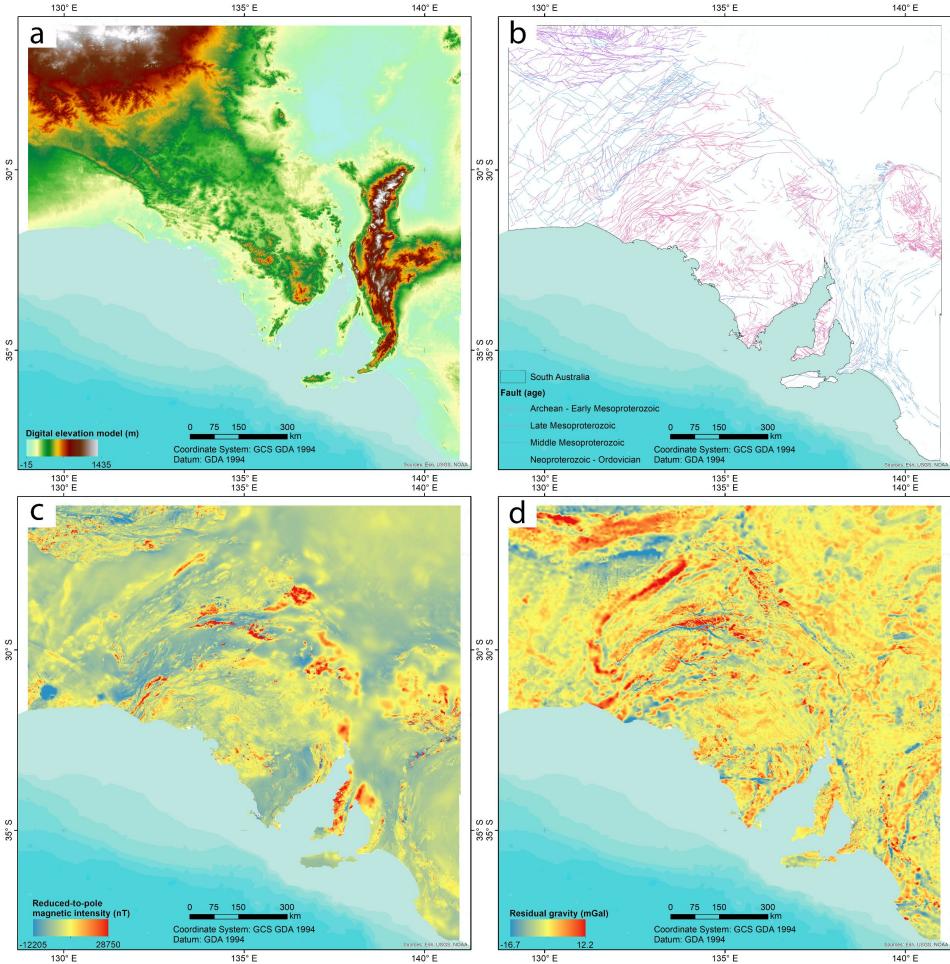




<b>MAPPABLE CRITERIA</b>	<ul style="list-style-type: none"> <li>Deformation</li> <li>Connate brines</li> <li>Enriched source rocks</li> <li>Structures</li> <li>Structures</li> <li>Deformation</li> <li>Metamorphism</li> <li>Magmatic fluids</li> <li>Permeable zones</li> <li>Chemical traps</li> <li>Aquifers</li> <li>Magmatism</li> <li>Meteoric fluids</li> <li>Magmatic fluids</li> </ul>
<b>SPATIAL PROXIES</b>	<ul style="list-style-type: none"> <li>Metamorphic grade</li> <li>Evaporites</li> <li>Radiometric anomalies</li> <li>Fault/shear zones</li> <li>Dilatational traps</li> <li>Magnetic/radiometric/geochemical anomalies</li> <li>Igneous intrusions</li> <li>Organics</li> <li>Folds</li> <li>Reactive rocks</li> <li>Alteration</li> <li>Alteration</li> <li>Sedimentary thickness</li> <li>Isotopes</li> <li>Geochemical anomalies</li> <li>Geophysical/geochemical anomalies</li> <li>Geophysical/geochemical anomalies</li> <li>Structures</li> </ul>

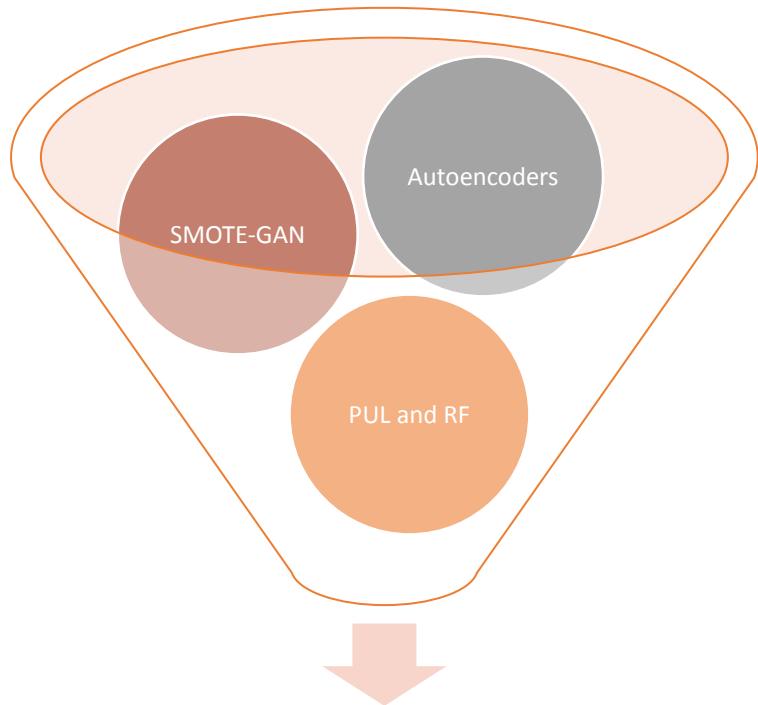
# Features

- A big volume of publicly available exploration datasets including geological, geochemical, and geophysical datasets are used to generate different features.
- Features highlighted by the model as important can aid in furthering the understanding of mineral systems and formation processes, and guide future exploration data collection requirements.



# Framework

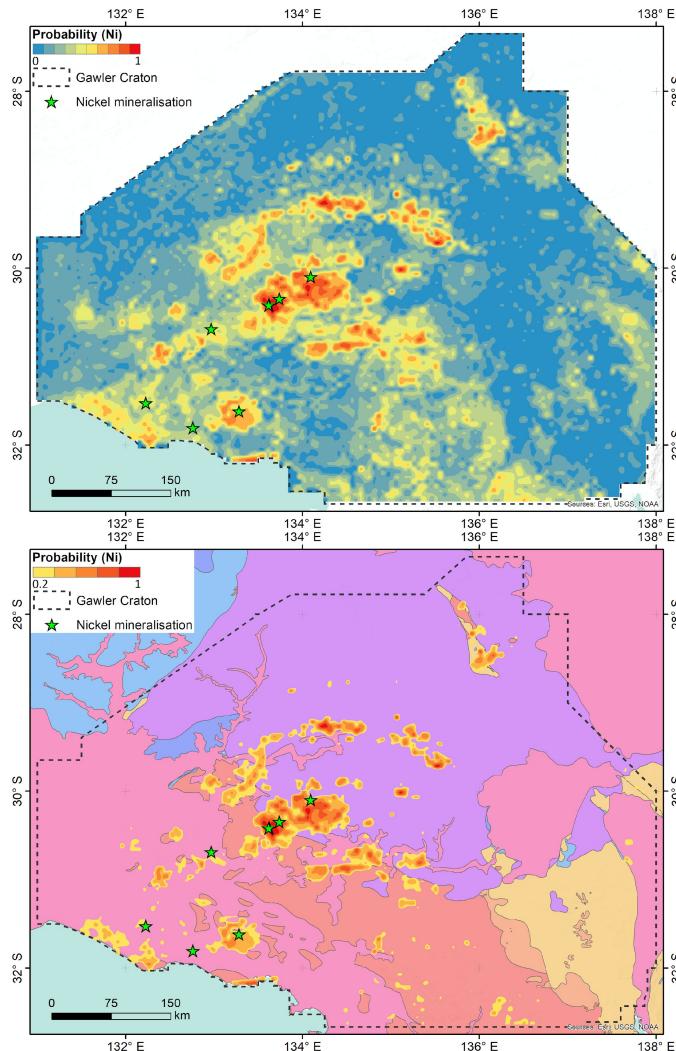
- The framework benefits from different machine learning methods for various purposes:
  - an improved generative adversarial network to overcome the class imbalance problem of the training dataset;
  - an autoencoder neural network to reduce the dimension of feature space;
  - the combination of a positive and unlabeled learning method and random forest as the main classifier for predicting probability values at target points.



Mineral Prospectivity Map

# Prospectivity Map

- Based on the results and different evaluation metrics, the model's performance is stable, and its accuracy is higher than the map generated by a conventional approach using a standard random forest classifier.
- The prospectivity map shows a strong spatial correlation between high probability values and known mineral occurrences and predicts a number of potential greenfield regions with as yet undiscovered deposits.

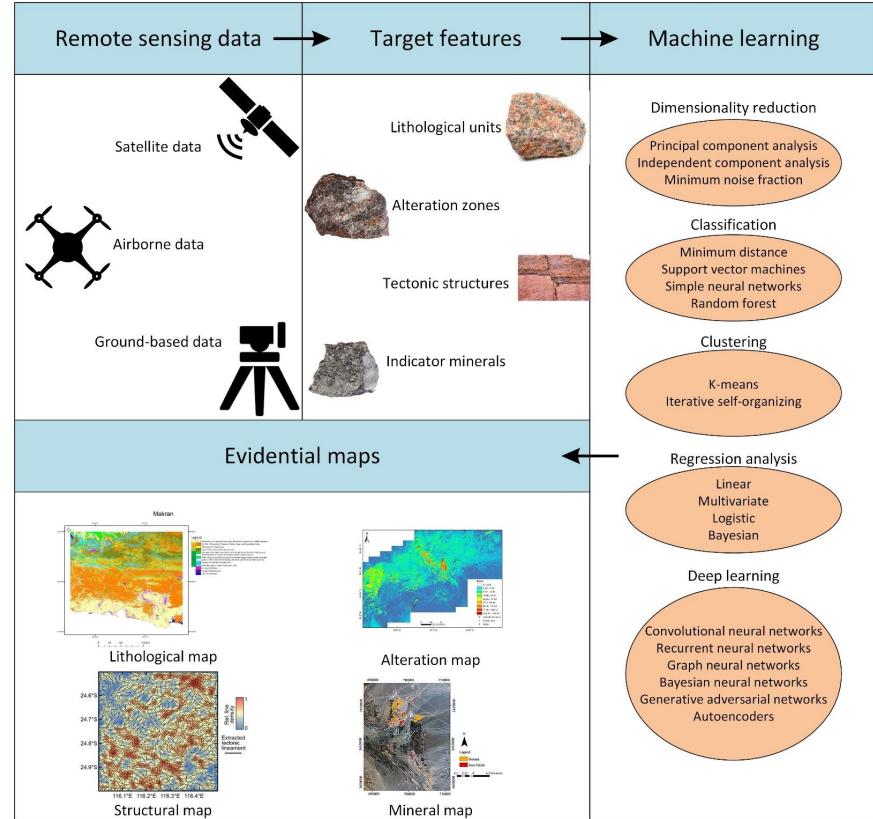


# Autoencoders and Remote Sensing for mineral exploration

Ehsan Farahbakhsh

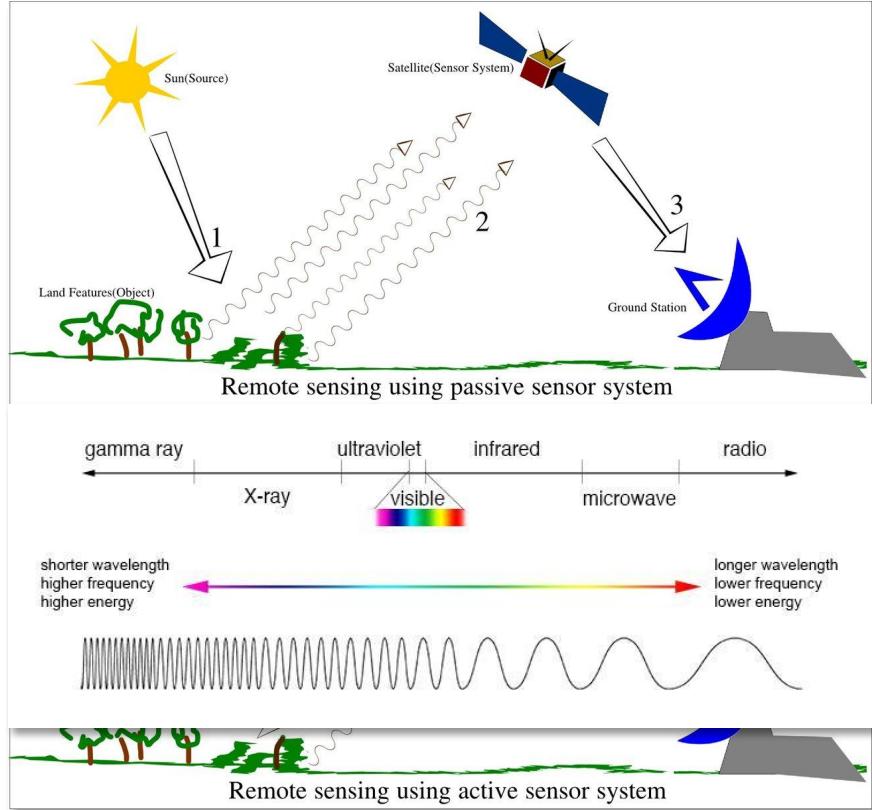


Rohitash Chandra

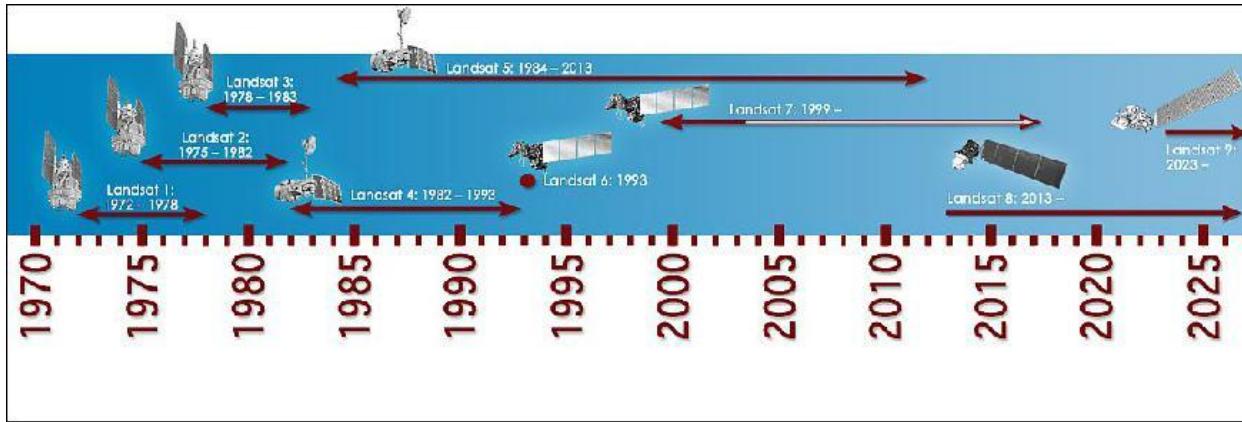


# Remote Sensing

- Passive sensors gather radiation that is emitted or reflected by the object or surrounding areas. Reflected sunlight is the most common source of radiation measured by passive sensors.
- Active collection, on the other hand, emits energy in order to scan objects and areas whereupon a sensor then detects and measures the radiation that is reflected or backscattered from the target.

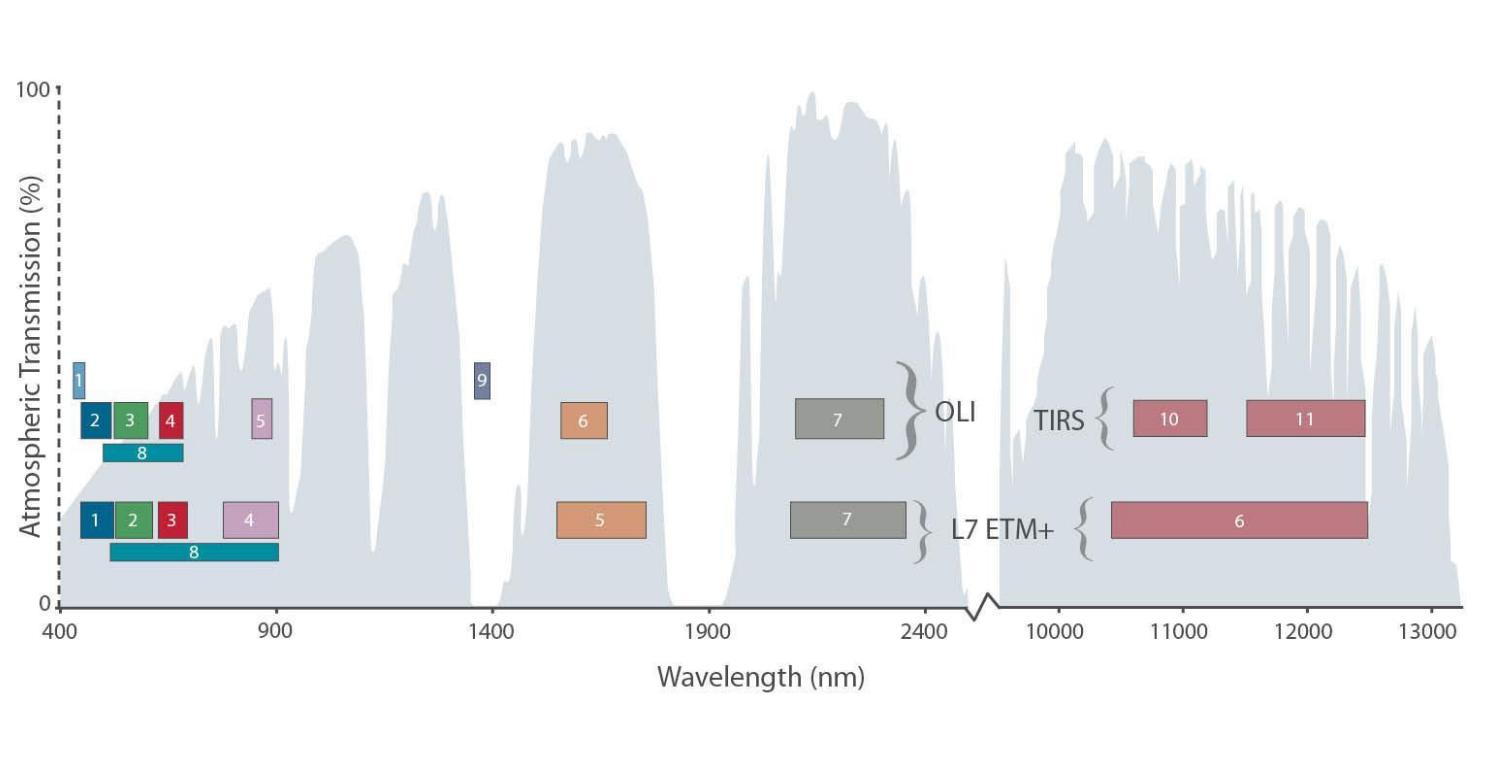


# Landsat

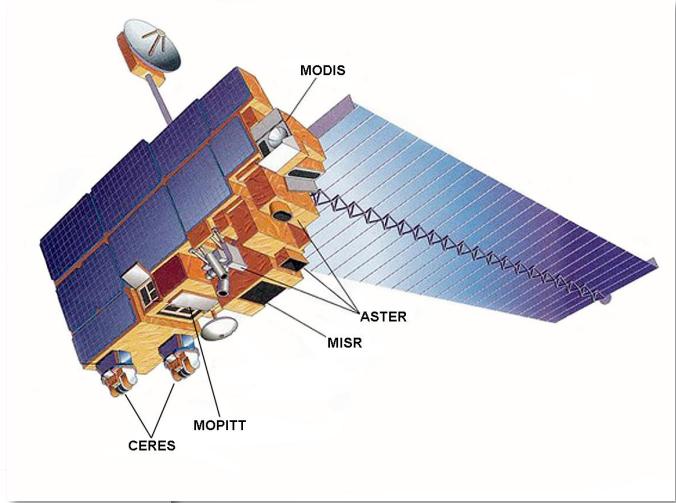
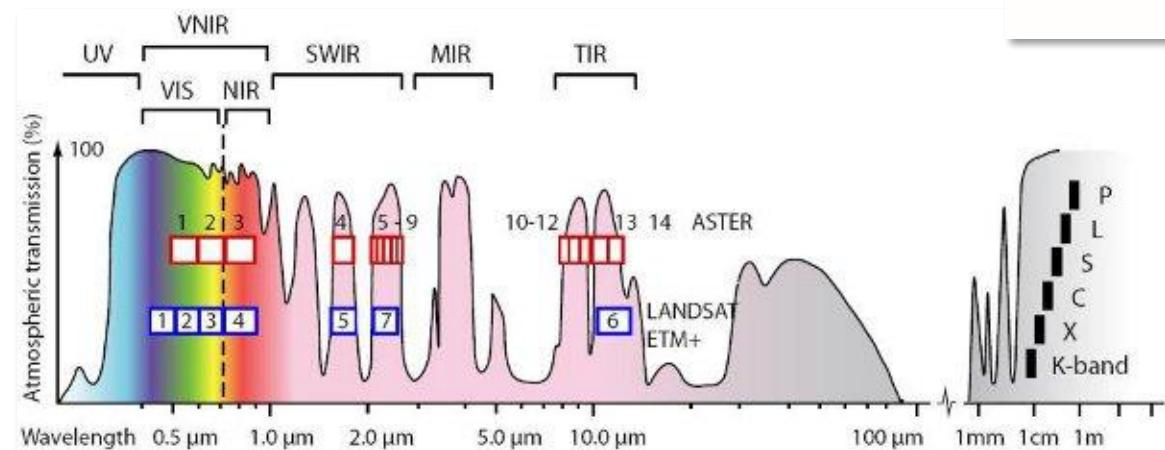


Satellite	Sensor	Swath	Bits	VNIR	SWIR	TIR
L8	OLI	185km	12	30m 30m 15m	30m	30m
	TIRS					100m 100m
Landsat 7	ETM+	185km	8	30m 30m 15m	30m 30m	60m
Landsat 4 & 5	MSS	185km	8	82m 30m 30m 30m	82m 30m 30m	120m
Landsat 1-2	RBV	183km		80m 80m 80m		
Landsat 3	RBV	183km		40m		
Landsat 1-3	MSS	183km	8	79m 79m 79m 79m		240m (L3 Only)

# Landsat 8

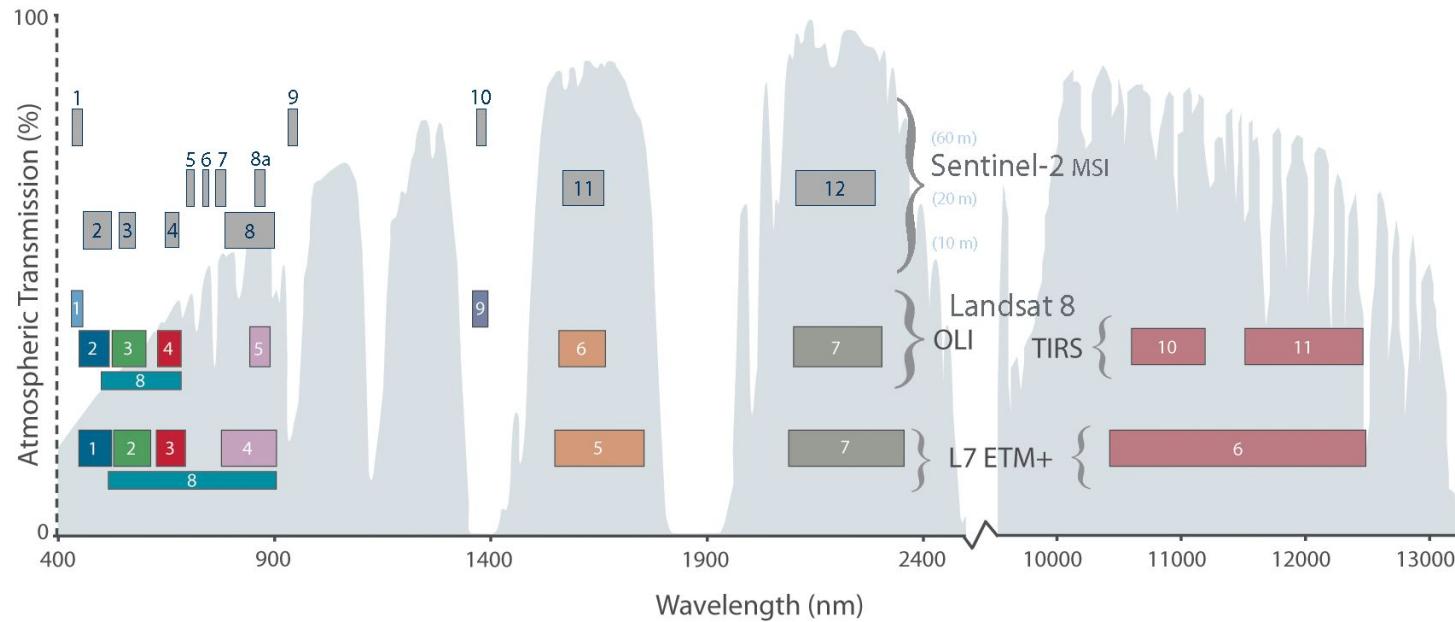


# ASTER

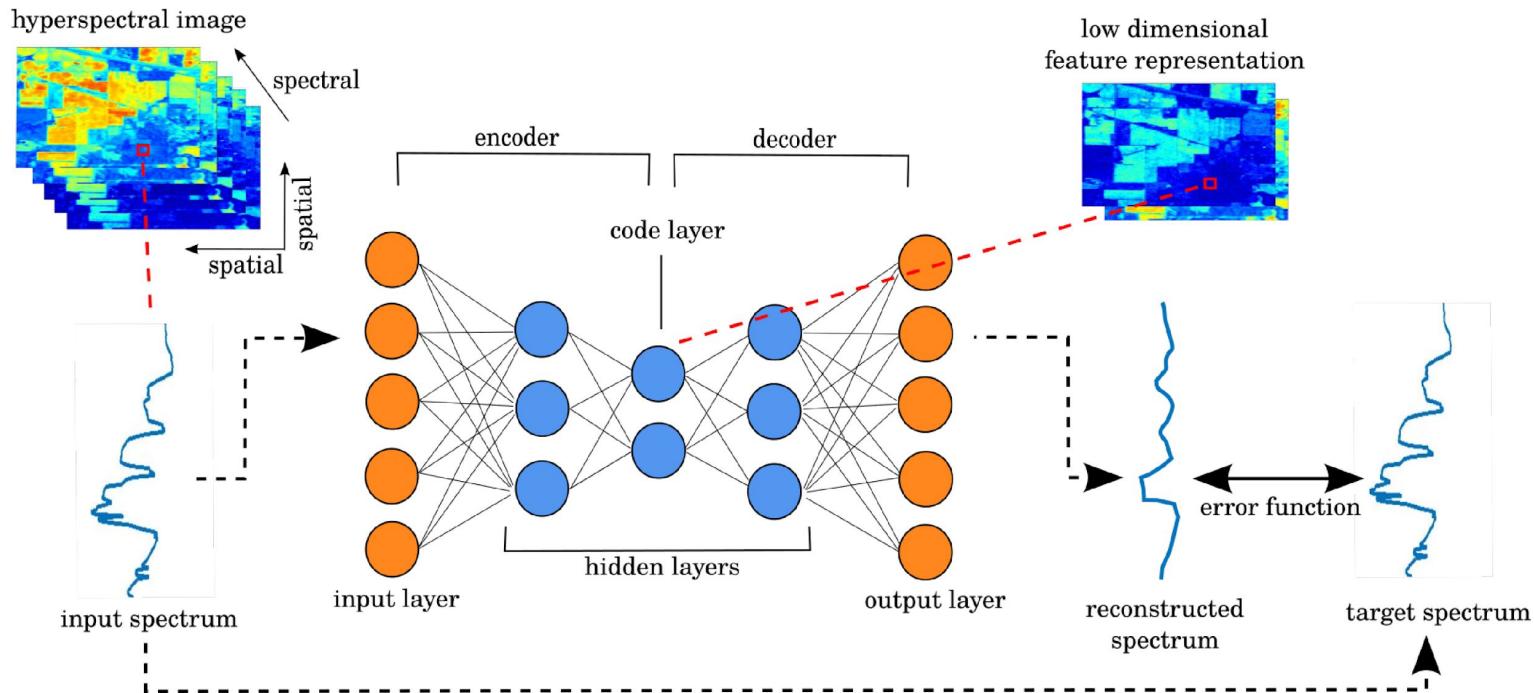


# Sentinel-2

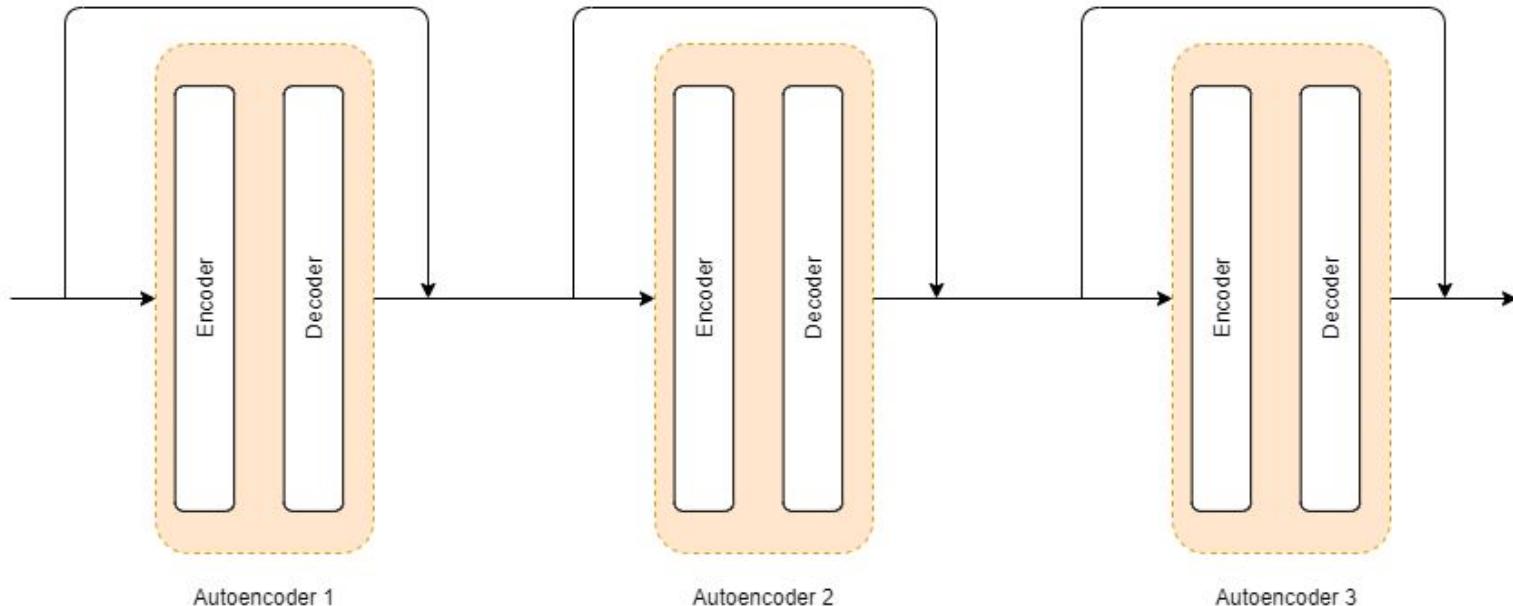
Comparison of Landsat 7 and 8 bands with Sentinel-2

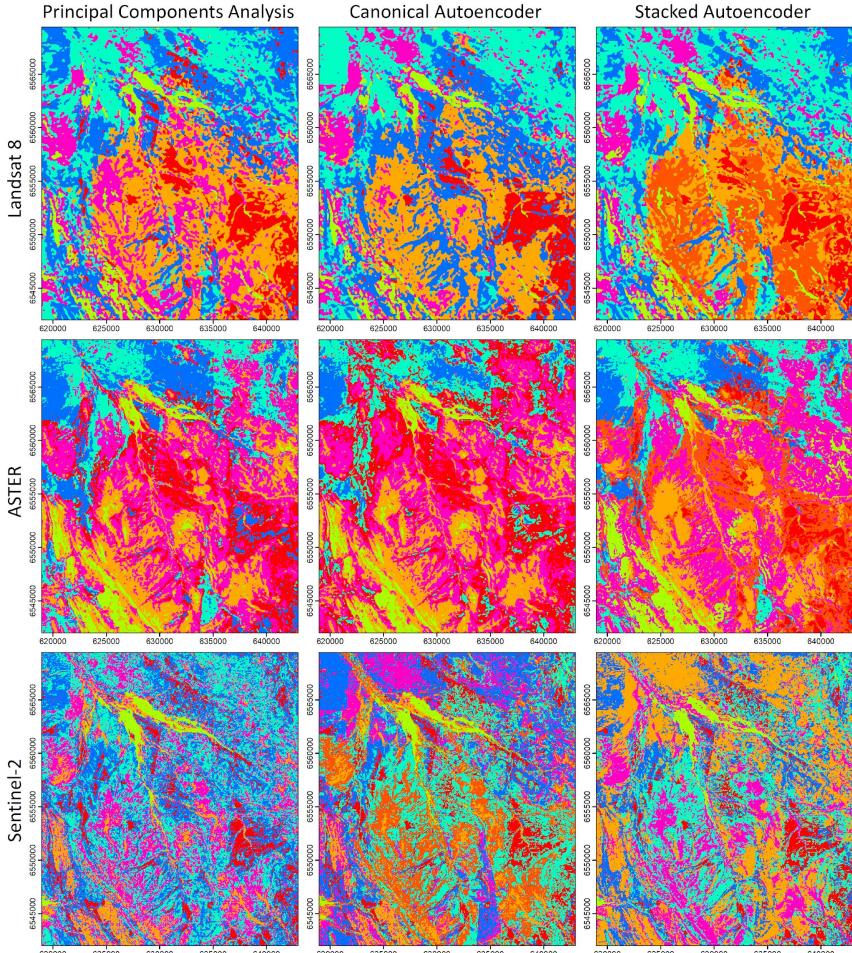
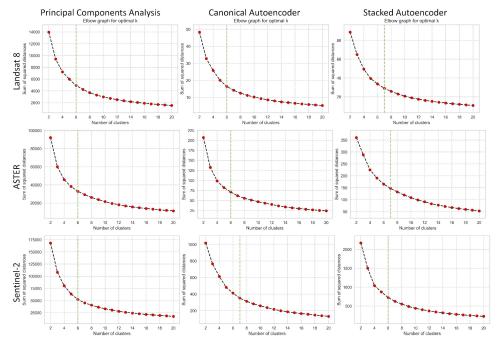
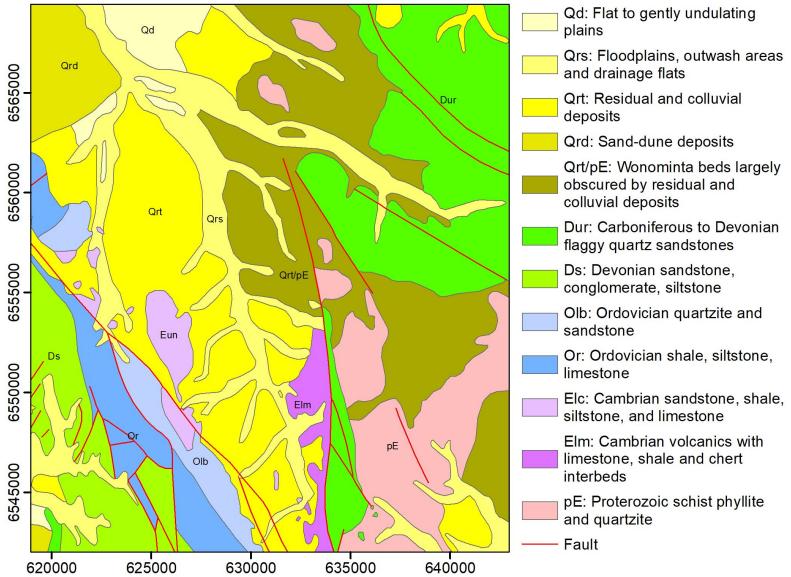


# Autoencoders



# Stacked Autoencoders







Prof. Dietmar Muller  
Former ARC Laureate Fellow,  
EarthByte Group, University of  
Sydney

# Coupling deep learning with Plate Tectonics models for mineral exploration



Julian Diaz-Rodriguez  
Master of Science, University of Sydney

J Diaz-Rodriguez, RD Muller, **R Chandra**. Predicting the emplacement of Cordilleran porphyry copper systems using a spatio-temporal machine learning model. *Ore Geology Reviews* 137, 2021: 104300



**EarthBYTE**  
Building a Virtual Earth

# Introduction

Porphyry copper (Cu) systems occur along magmatic belts derived in subduction zones. Our current understanding of their formation is restricted to observations from the overriding plate, resulting in a knowledge gap in terms of processes occurring in convergence zones through time.

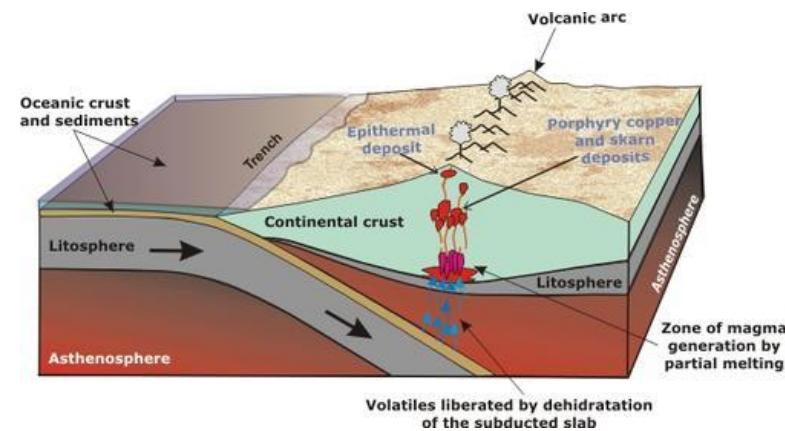
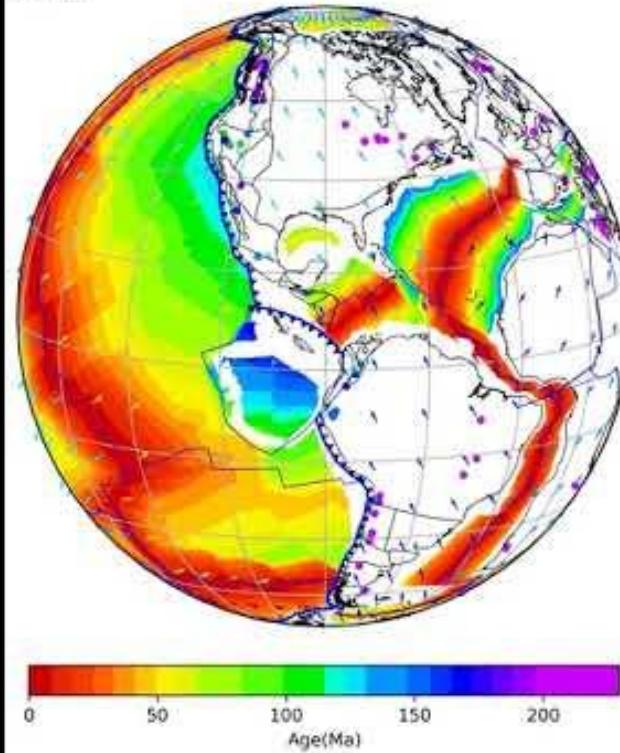
An association between key tectonic processes and the timing and location of porphyry Cu systems requires linking geological observations to plate tectonic subduction models.

We connect the evolution of subduction zones and downgoing slab properties with the history of porphyry ore deposition across the Americas by using a spatio-temporal machine learning approach.

We use these spatio-temporal properties to apply a wide range of prominent machine learning methods and show the results in terms of accuracy of predictions on the test dataset.

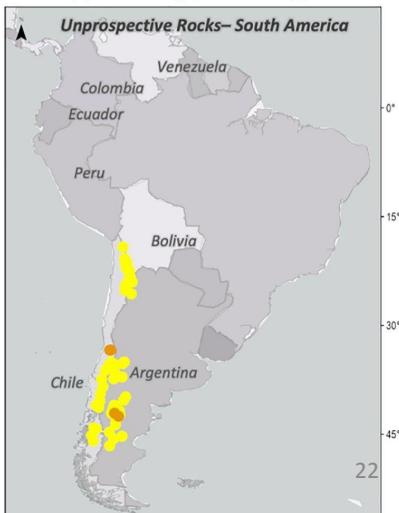
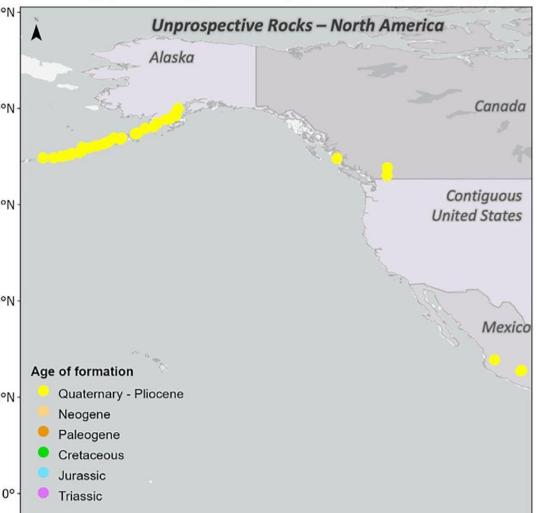
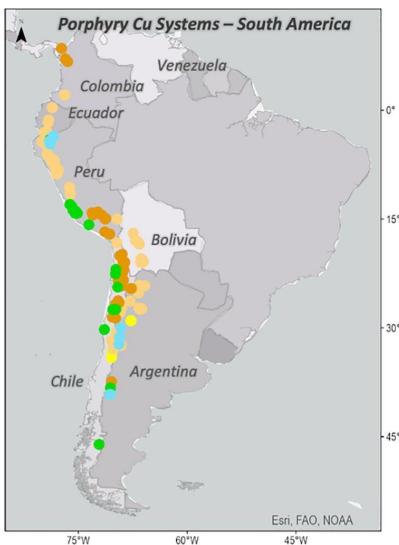
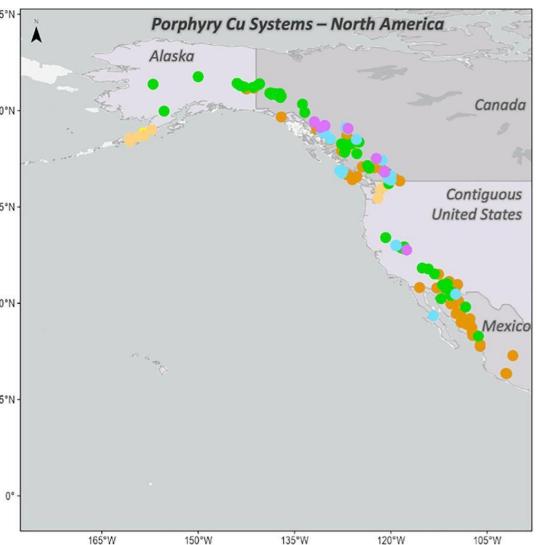
We incorporate plate tectonic subduction models with machine learning models for the generation of copper mineralization prediction maps.

96 Ma



Schematic diagram of the tectonic setting of porphyry, epithermal and skarn deposits.  
Source:

[https://www.geo.fu-berlin.de/en/v/geolearning/mountain\\_building/resources/porphyry\\_copper\\_ore\\_deposits/index.html](https://www.geo.fu-berlin.de/en/v/geolearning/mountain_building/resources/porphyry_copper_ore_deposits/index.html)



# GPlates

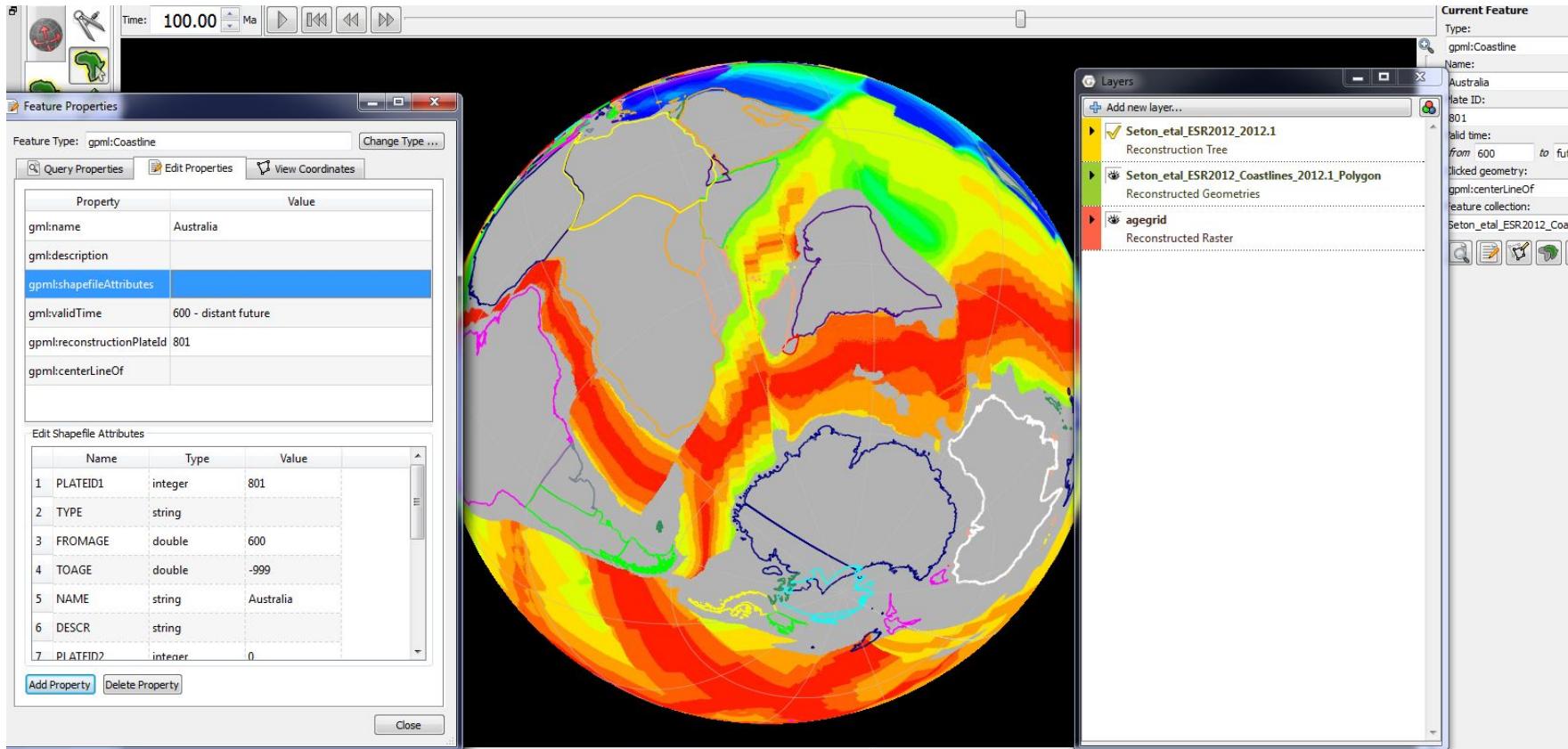
The GPlates homepage features a large, semi-transparent globe visualization in the background. In the foreground, the title "GPlates" is displayed in a large, bold, dark font. Below the title is a descriptive paragraph: "GPlates is a plate tectonics program. Manipulate reconstructions of geological and paleogeographic features through geological time. Interactively visualize vector, raster and volume data. PyGPlates is the GPlates Python library. Get fine-grained access to GPlates functionality in your Python scripts." To the left of the text, there are four small circular thumbnails showing different global geological reconstructions. Below these thumbnails is the text "Latest release v2.2". A green button labeled "Download Now" with a download icon is positioned below the thumbnails. To the right of the text, a large, detailed globe visualization shows geological features like plate boundaries, mountain ranges, and ocean floor topography in various colors (blue, red, yellow, green). The globe is overlaid with numerous small colored dots and arrows, likely representing data points and movement vectors.



EarthBYTE

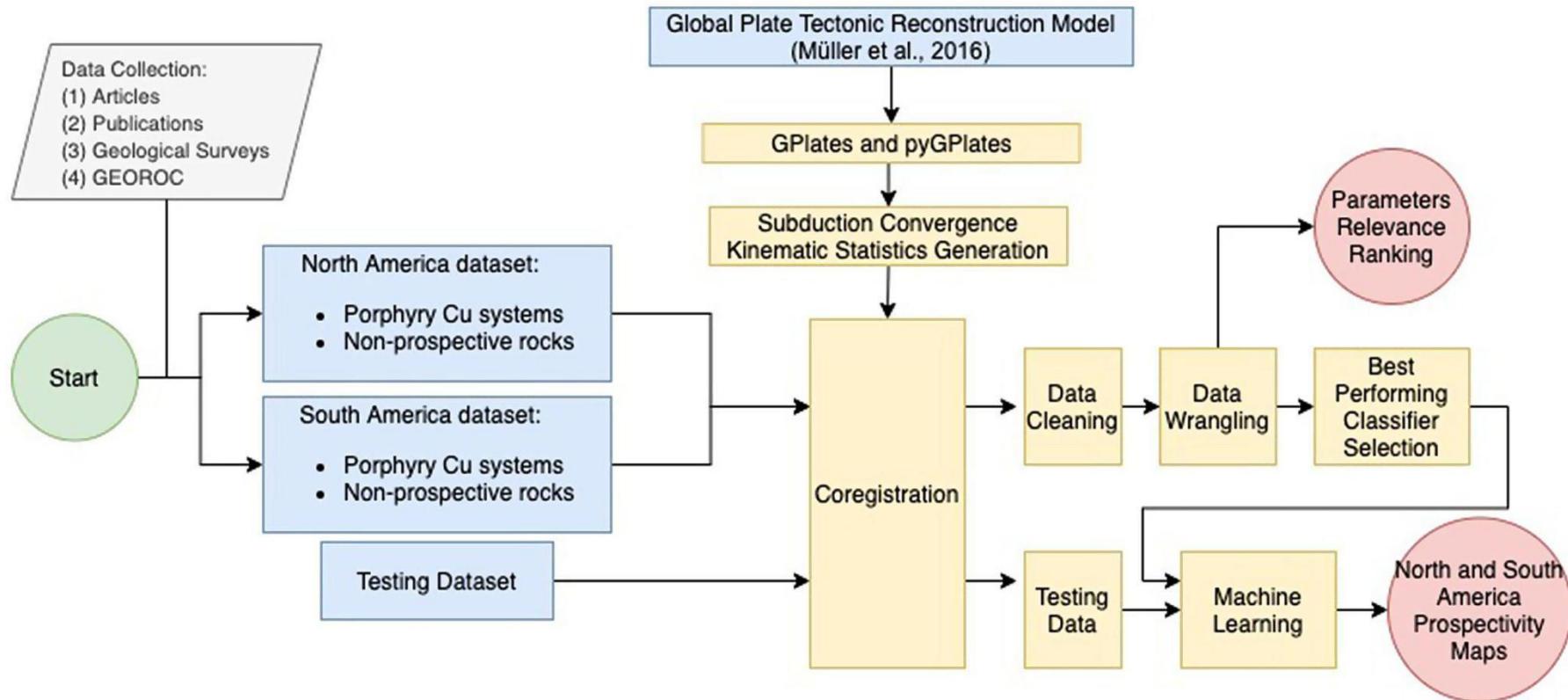
Building a Virtual Earth

# GPlates Software

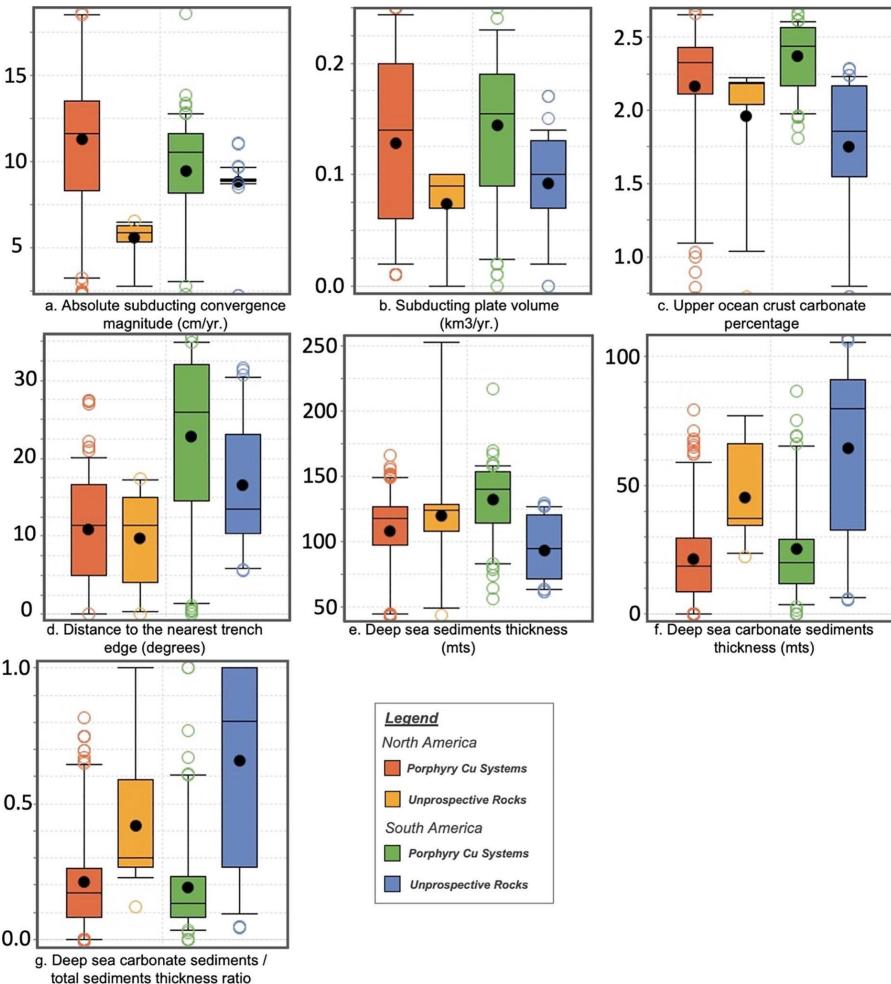


Source:

# Framework: Coupling Plate Tectonics with Machine Learning



# Data Visualization



Mean values and standard deviation of selected parameters related to the dynamic evolution of subduction zones at the time of formation of porphyry Cu systems and non-prospective rocks in North America and South America. \* Kilometres calculated using a great circle distance formula for comparison purposes (in B&W).

Parameters	North America			South America				
	Porphyry Cu systems		Unprospective Rocks	Porphyry Cu systems		Unprospective Rocks	Mean	SD
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Subducting convergence magnitude – Absolute vector (cm/yr.)	11.30	4.87	5.59	1.03	9.46	3.16	8.84	1.08
Subducting plate volume (km <sup>3</sup> /yr.)	0.13	0.07	0.07	0.03	0.14	0.06	0.09	0.04
Upper ocean crust carbonate percentage	2.17	0.47	1.96	0.42	2.37	0.22	1.75	0.42
Distance to the nearest trench edge (degrees) (kilometres*)	10.88 (1210)	7.01 (780)	9.74 (1083)	5.72 (636)	22.79 (2534)	10.65 (1184)	16.56 (1841)	8.04 (894)
Deep sea total sediments thickness (m)	108.30	31.25	120.15	42.92	132.43	26.09	93.49	23.00
Deep sea carbonate sediments thickness (m)	21.23	17.06	45.24	16.63	25.21	30.38	64.39	32.95
Deep sea carbonate sediment / total sediment thickness ratio	0.21	0.19	0.42	0.22	0.19	0.18	0.66	0.36

Correlation matrix using Spearman's correlation score for the dynamic evolution parameters related to the formation of porphyry Cu systems for South America. Strong correlations are defined as Spearman's correlation scores above 0.75. Grey colours are for scores below 0.75. Red shades of colour range between 0.75 and 1.0. Blue colours are for the main diagonal of the matrix. (in colour).

South America	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Subd. Conv. Magn. - Absolute Vect.	1.00	<b>0.91</b>	-0.01	0.48	-0.38	-0.15	-0.88	<b>0.83</b>	-0.06	0.17	-0.11	0.06	<b>0.80</b>	0.11	0.12
2. Subd. Conv. Magn. - Orthogonal Vect.	<b>0.91</b>	1.00	-0.14	0.37	-0.48	-0.01	-0.83	<b>0.91</b>	-0.18	0.18	-0.10	0.04	<b>0.85</b>	0.03	0.03
3. Subd. Conv. Magn. - Parallel Vect.	-0.01	-0.14	1.00	0.26	0.16	-0.69	0.09	-0.10	<b>0.96</b>	0.03	0.02	0.12	-0.06	0.20	0.18
4. Trench Plate Magn. - Absolute Vect.	0.48	0.37	0.26	1.00	-0.75	-0.23	-0.12	0.11	0.29	0.42	-0.12	0.44	0.49	0.56	0.58
5. Trench Plate Magn. - Orthogonal Vect.	-0.38	-0.48	0.16	-0.75	1.00	-0.19	0.10	-0.17	0.10	-0.48	0.09	-0.34	-0.54	-0.43	-0.46
6. Trench Plate Magn. - Parallel Vect.	-0.15	-0.01	-0.69	-0.23	-0.19	1.00	0.13	-0.12	-0.48	-0.03	0.07	-0.09	-0.07	-0.15	-0.10
7. Down-going Plate Magn. - Absolute Vect.	-0.88	-0.83	0.09	-0.12	0.10	0.13	1.00	-0.92	0.16	-0.01	0.15	0.08	-0.70	0.08	0.08
8. Down-going Plate Magn. - Orthogonal Vect.	<b>0.83</b>	<b>0.91</b>	-0.10	0.11	-0.17	-0.12	-0.92	1.00	-0.18	0.04	-0.16	-0.07	<b>0.77</b>	-0.10	-0.13
9. Down-going Plate Magn. - Parallel Vect.	-0.06	-0.18	<b>0.96</b>	0.29	0.10	-0.48	0.16	-0.18	1.00	0.06	0.06	0.17	-0.08	0.23	0.24
10. Distance to the nearest trench edge	0.17	0.18	0.03	0.42	-0.48	-0.03	-0.01	0.04	0.06	1.00	-0.24	0.58	0.45	0.62	0.61
11. Deep-sea carbonate sediment thickness	-0.11	-0.10	0.02	-0.12	0.09	0.07	0.15	-0.16	0.06	-0.24	1.00	-0.31	-0.23	-0.19	-0.17
12. Upper-ocean crust carbonate percentage	0.06	0.04	0.12	0.44	-0.34	-0.09	0.08	-0.07	0.17	0.58	-0.31	1.00	0.36	<b>0.82</b>	<b>0.81</b>
13. Subducting plate volume	<b>0.80</b>	<b>0.85</b>	-0.06	0.49	-0.54	-0.07	-0.70	<b>0.77</b>	-0.08	0.45	-0.23	0.36	1.00	0.40	0.39
14. Sea floor age	0.11	0.03	0.20	0.56	-0.43	-0.15	0.08	-0.10	0.23	0.62	-0.19	<b>0.82</b>	0.40	1.00	<b>0.99</b>
15. Deep-sea sediment total thickness	0.12	0.03	0.18	0.58	-0.46	-0.10	0.08	-0.13	0.24	0.61	-0.17	<b>0.81</b>	0.39	<b>0.99</b>	1.00

# Results

Cross validation results for different supervised machine learning classifiers. Cross validation calculated using a ten-fold with a shuffle split on 80/20 training/ testing ratio. Range of colours just for indicative purposes from Gray (up to 85%) and shades of yellow (85%) to red (100%). The complete data used for cross validation is included in supplementary data. (in colour).

	North America		South America	
	Mean	SD	Mean	SD
Support Vector (RBF k.)	98.3%	1.7%	97.9%	2.0%
Random Forest	95.4%	4.3%	97.4%	1.7%
Multi-layer Perceptron	96.5%	3.9%	96.0%	1.6%
Gaussian Process	92.8%	3.7%	94.7%	3.3%

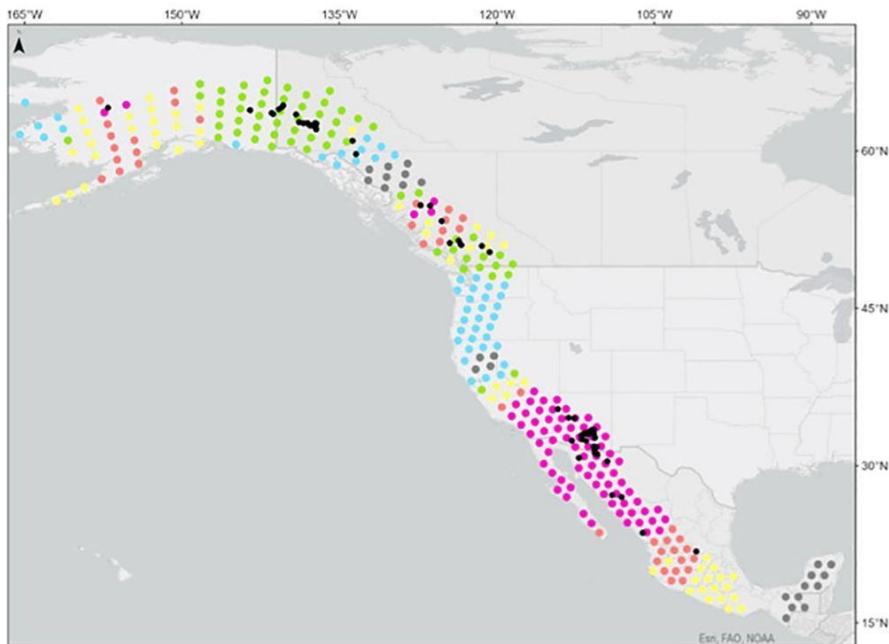
Confusion matrix performance metrics for different supervised machine learning classifiers for North America. (in colour).

Classifier	North America					Recall	Precision	Accuracy	F1	AUC score
	True Negative	False Positive	False Negative	True Positive						
Support Vector (RBF k.)	9	1	0	47		100.0%	97.9%	98.2%	98.9%	100.0%
Multi-layer Perceptron	9	1	0	47		100.0%	97.9%	98.2%	98.9%	100.0%
Gaussian Process	7	3	1	46		97.9%	93.9%	93.0%	95.8%	99.4%
Random Forest	9	1	2	45		95.7%	97.8%	94.7%	96.8%	98.7%

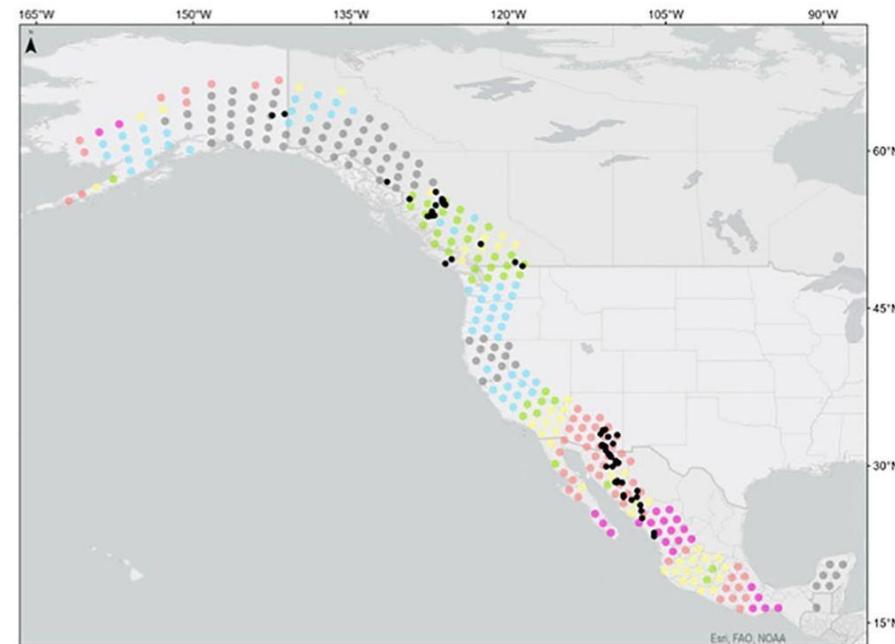
Confusion matrix performance metrics for different supervised machine learning classifiers for South America. (in colour).

Classifier	South America					Recall	Precision	Accuracy	F1	AUC score
	True Negative	False Positive	False Negative	True Positive						
Support Vector (RBF k.)	25	0	0	29		100.0%	100.0%	100.0%	100.0%	100.0%
Random Forest	24	1	1	28		96.6%	96.6%	96.3%	96.6%	99.2%
Multi-layer Perceptron	23	2	1	28		96.6%	93.3%	94.4%	94.9%	97.9%
Gaussian Process	23	2	2	27		93.1%	93.1%	92.6%	93.1%	95.7%

a. 80 – 60 Ma.



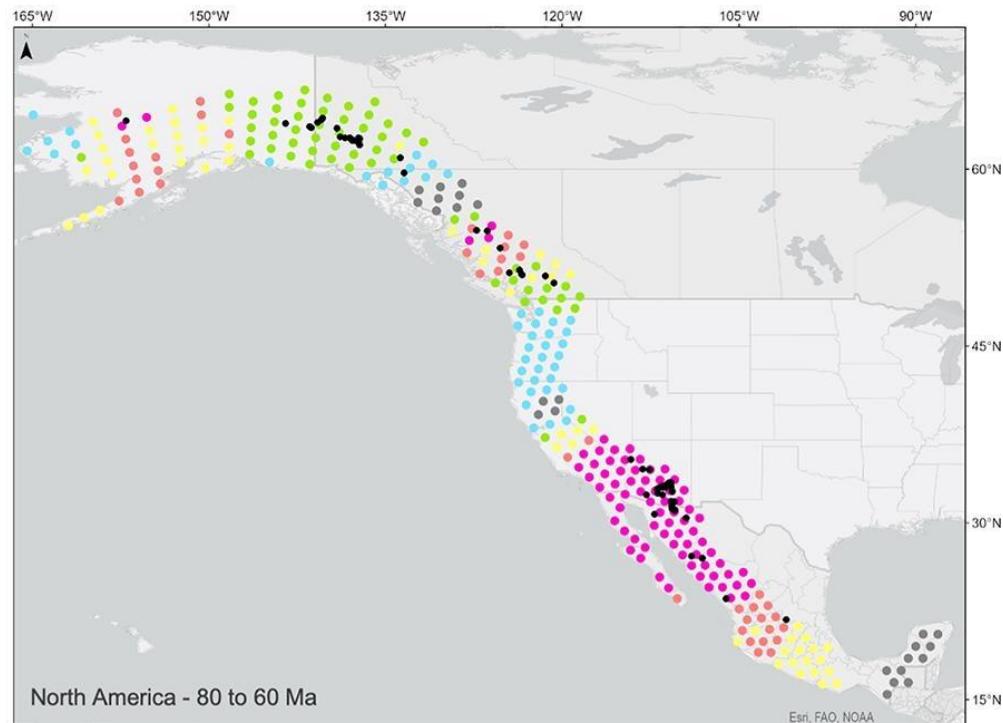
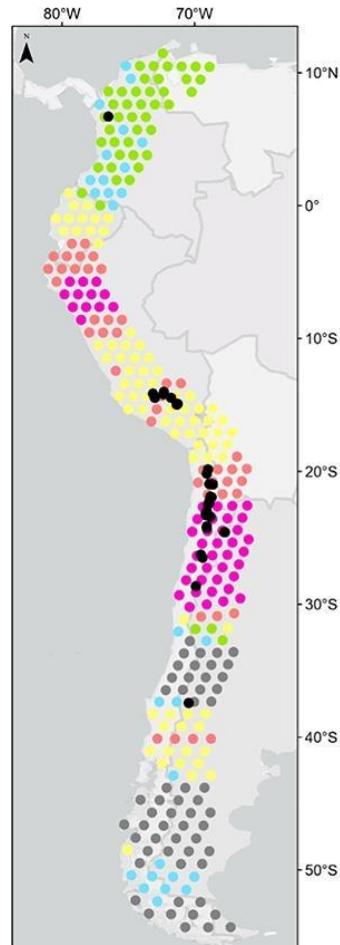
b. 60 – 40 Ma.

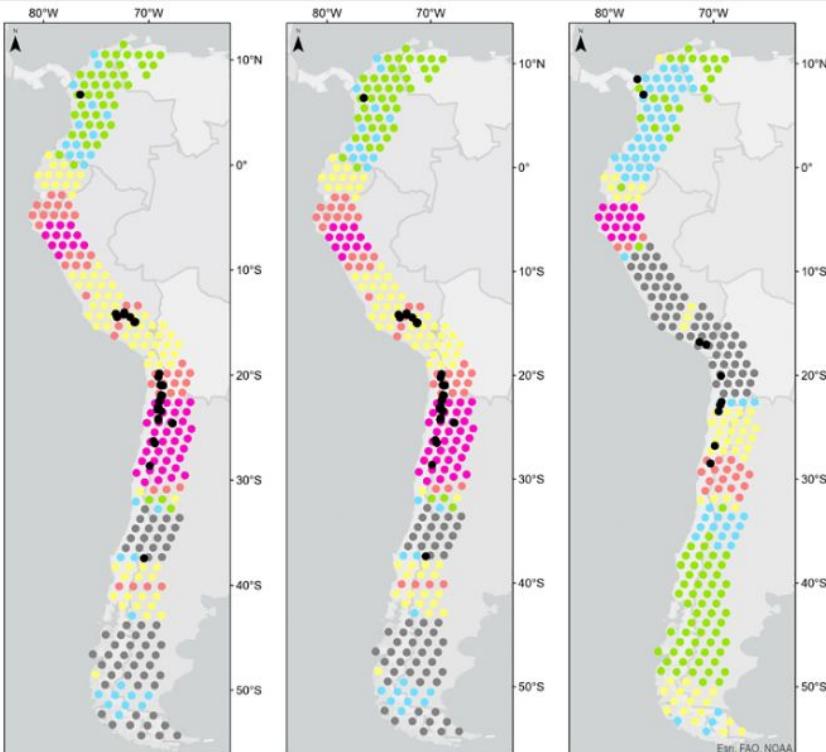


Legend

Score Prediction – Support Vector Classifier

- **Very High Prospectivity**
- **High Prospectivity**
- **High – Moderate Prospectivity**
- **Low – Moderate Prospectivity**
- **Low Prospectivity**
- **Very Low Prospectivity**
- **Porphyry Cu Systems**





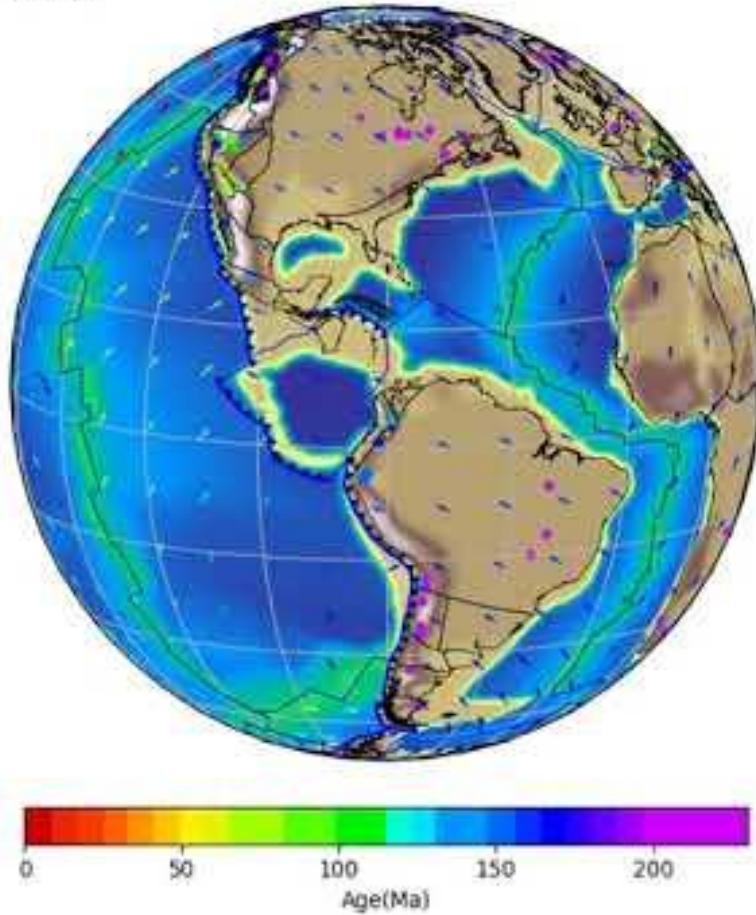
#### Legend

Score Prediction – Support Vector Classifier

- |                                 |                          |
|---------------------------------|--------------------------|
| • Very High Prospectivity       | ● Low Prospectivity      |
| • High Prospectivity            | ● Very Low Prospectivity |
| • High – Moderate Prospectivity | ● Porphyry Cu Systems    |
| • Low – Moderate Prospectivity  |                          |

Fig. 6. Prospectivity map for porphyry Cu systems in South America using a support vector classifier with RBF kernel classifiers. (a) 23 to 3 Ma. (b) 47 to 27 Ma. (c) 66 to 47 Ma. (in colour).

68 Ma



# Discussion

We propose a deep time spatio-temporal machine learning model to identify highly prospective areas and endowment ages for porphyry Cu mineralisation along the western Cordillera of North America and the Andes in South America.

This open-source workflow is based on pyGPlates which allows us to reconstruct in time and space the kinematic processes occurring at the boundary of convergent plates, including features of the downgoing plate such as deep sea sediment thickness and oceanic crustal or lithospheric properties. These parameters constrain the necessary geological processes to generate porphyry Cu systems under adequate overriding plate conditions.

We conclude that the most important of those parameters linked to the formation of porphyry Cu systems across North and South America is the absolute magnitude of convergence velocity. This magnitude is on average faster at the time when these systems formed as opposed to the rates related to the emplacement of non-prospective intrusions.

# Machine learning for drill-core analysis - mining



A/Prof. Stuart Clark  
UNSW



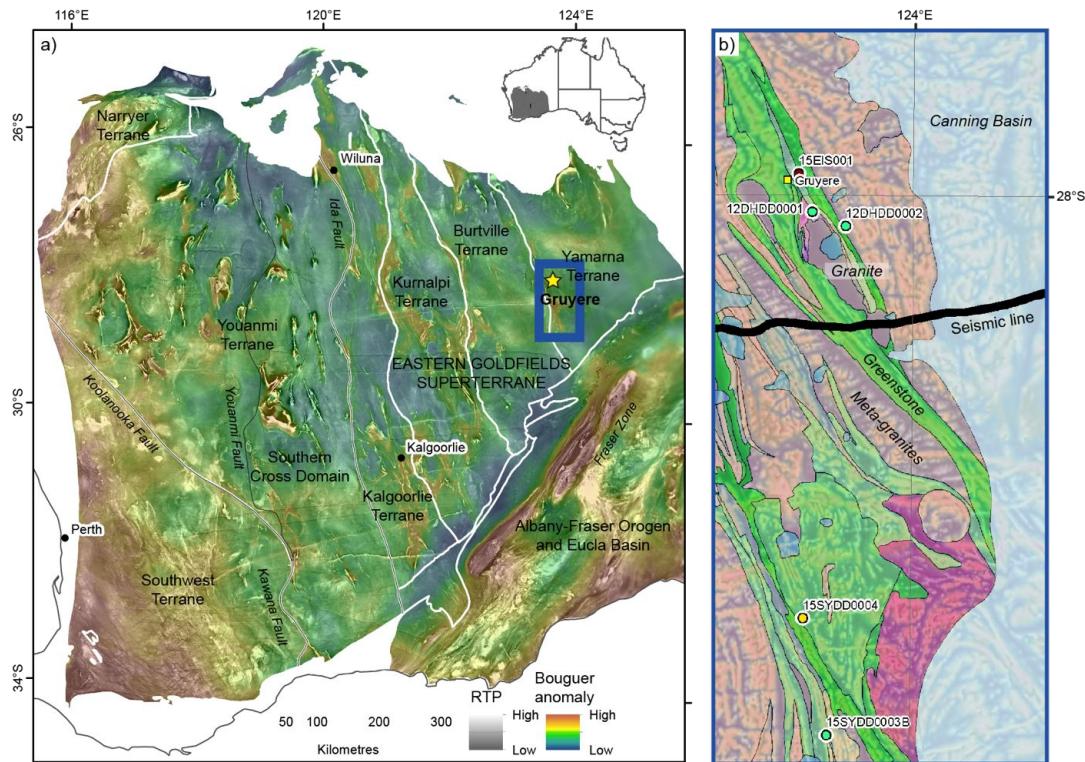
Subhash Chandra  
Former PhD student, UNSW



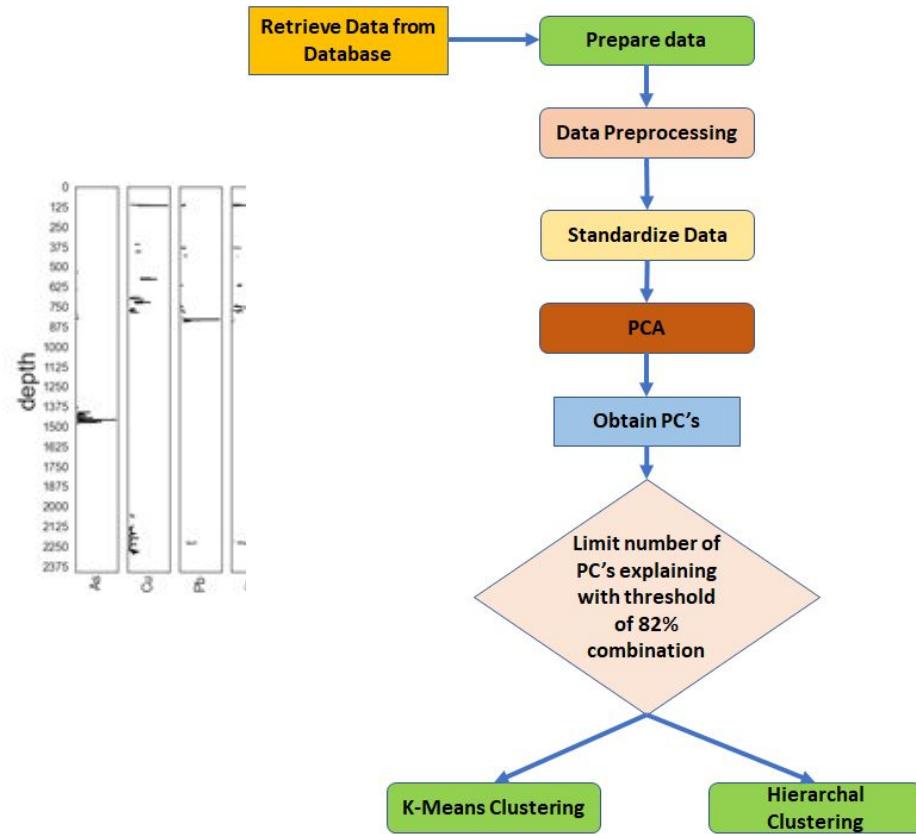
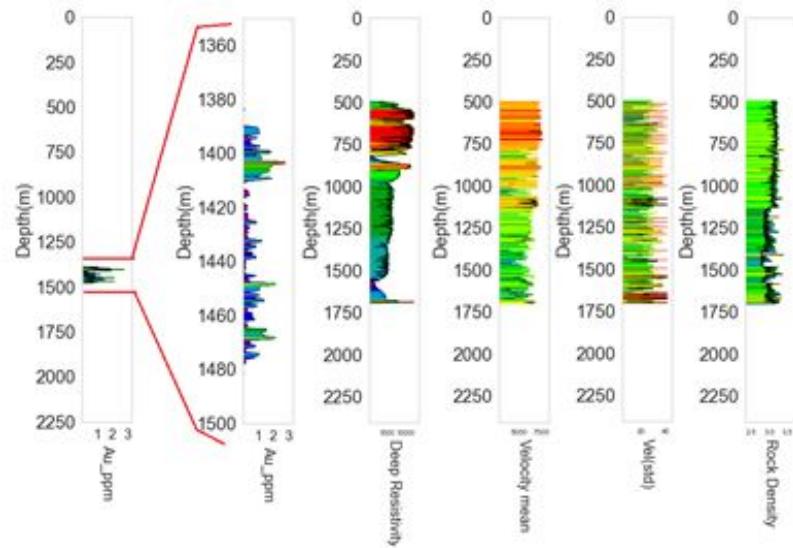
Dr Mark Lindsey  
CSIRO Science Lead -Mining



Ehsan  
Farahbakhsh,  
University of  
Sydney



S. Chandra, M. Lindsay, S. Clark, and R. Chandra, Drill-core analysis for mineral composition with an unsupervised machine learning framework, Ore Geology Reviews, 2022 (In review)



# Convolutional Neural Networks and Remote Sensing Lithological Mapping

H. Shirmard, E. Farahbakhsh, E. Heidari, A. B. Pour, B. Pradhan, R. D. Müller and **R. Chandra**. A comparative study of convolutional neural networks and conventional machine learning models for lithological mapping using remote sensing data. *Remote Sensing* 10, 1–22, January 2018.



Associate Professor  
Universiti Malaysia  
Terengganu, Malaysia



Prof. Biswajeet Pradhan,  
UTS Sydney



Hodjat Srimard, Exploration  
Geologist, Iran



Ehsan Farahbakhsh,  
Postdoctoral  
Research Associate,  
EarthByte Group,  
University of Sydney



Prof. Dietmar Muller  
Former ARC Laureate  
Fellow, EarthByte Group,  
University of Sydney

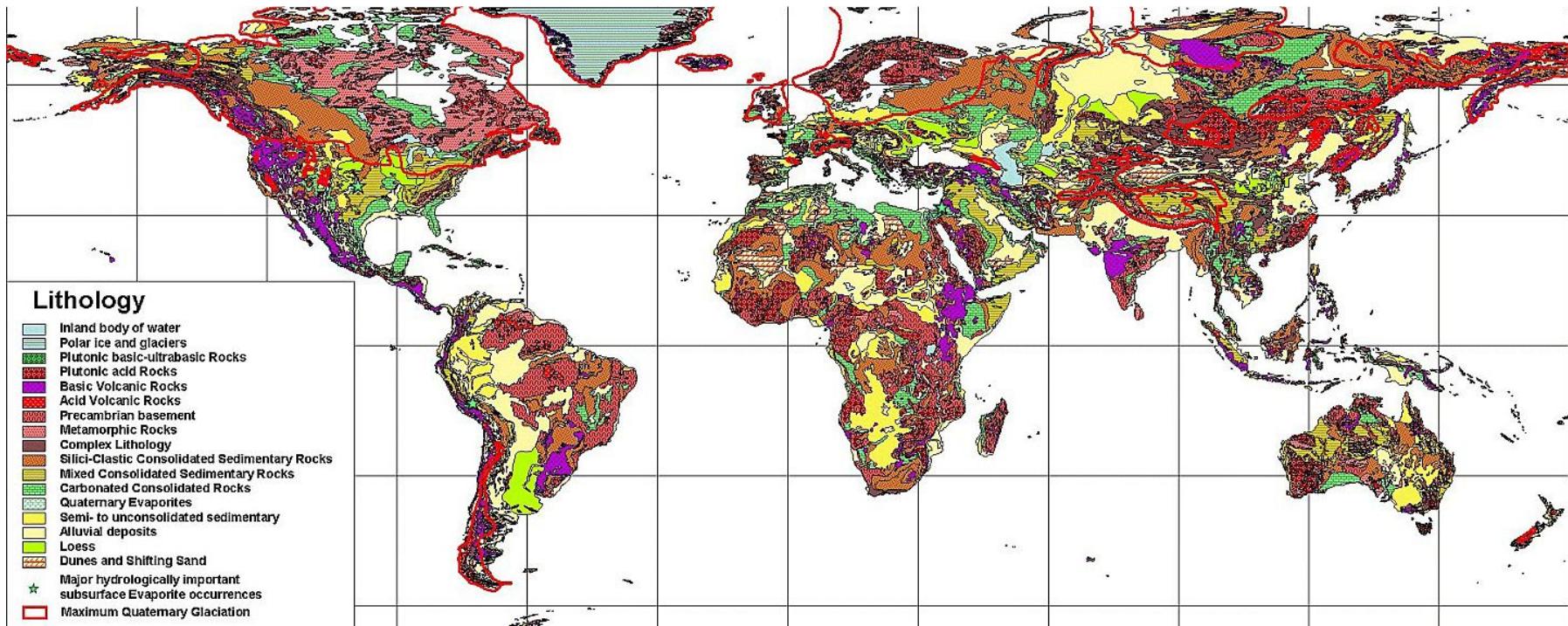
# Introduction

Lithological mapping is a critical aspect of geological mapping that can be useful in studying the mineralization potential of a region and has implications for mineral prospectivity mapping. This is a challenging task if done manually, particularly in highly remote areas that necessitates a large number of participants and resources.

The combination of machine learning methods and remote sensing data can provide an easy, low-cost, and accurate approach for mapping lithological units. In this study, we use deep learning via convolutional neural networks (CNNs) and conventional machine learning methods involving support vector machines and multilayer perceptron to map lithological units of a mineral-rich area in the southeast of Iran.

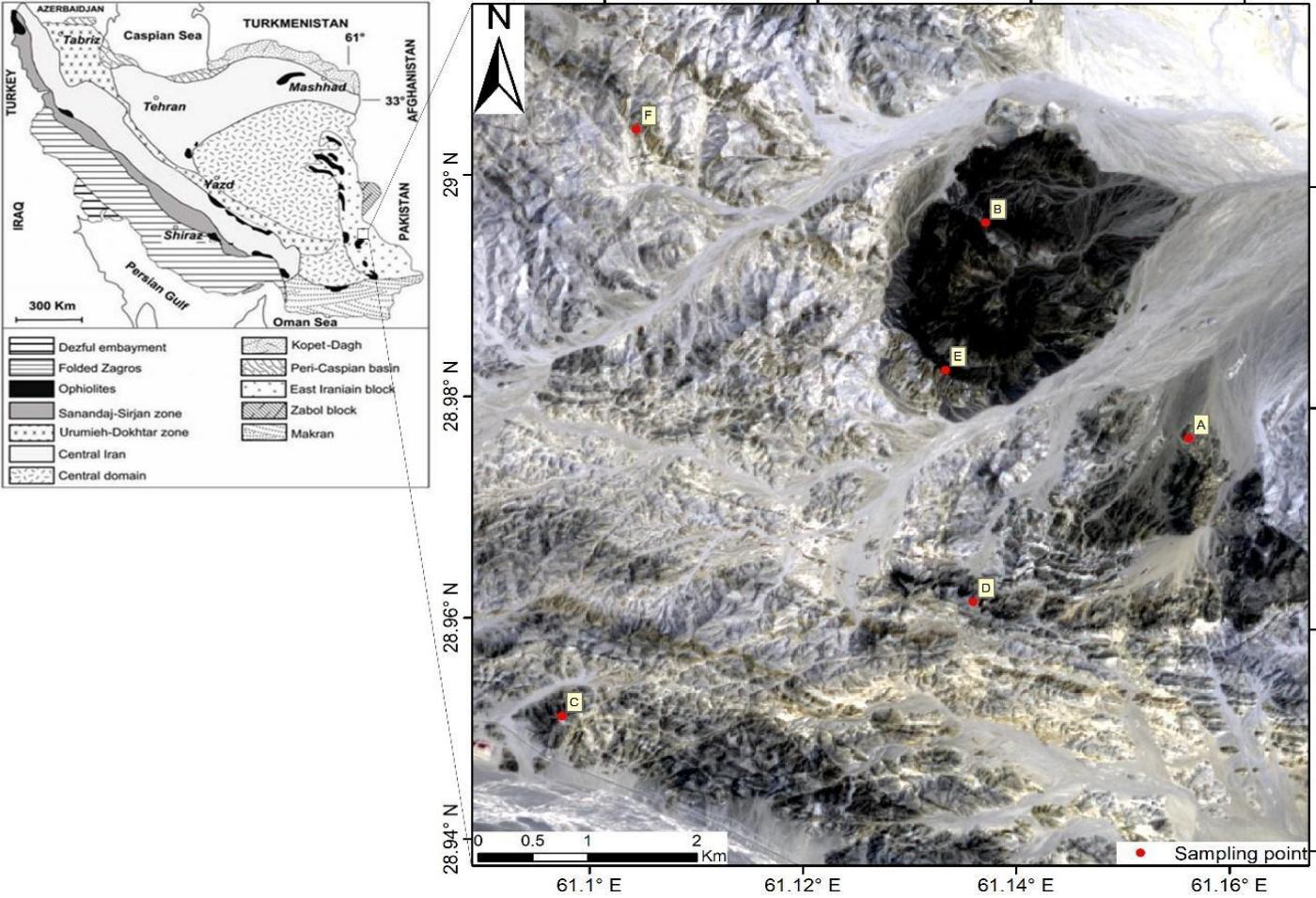
Moreover, we use and compare the efficiency of three different types of multispectral remote sensing data including operational land imager (OLI), advanced spaceborne thermal emission and reflection radiometer (ASTER), and Sentinel-2.

# Global lithology map



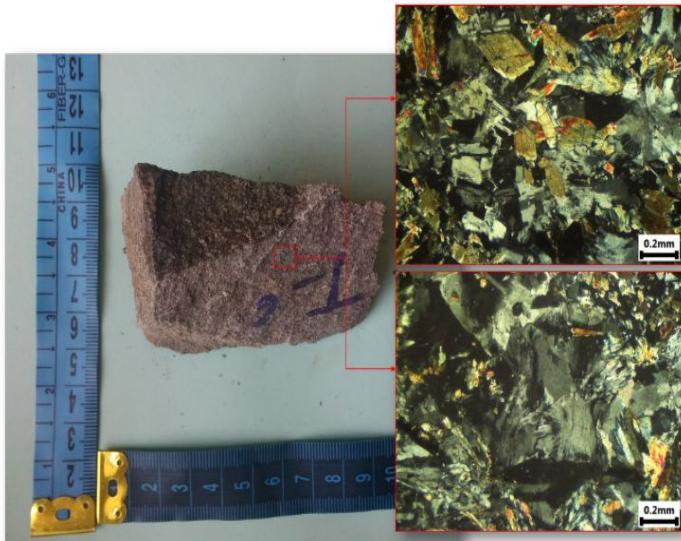
Dürr, H. H., Meybeck, M., & Dürr, S. H. (2005). Lithologic composition of the Earth's continental surfaces derived from a new digital map emphasizing riverine material transfer. *Global Biogeochemical Cycles*, 19(4).

**Figure:** Simplified tectonic map of Iran on the left and a true color image of the study area obtained by Sentinel-2 data on the right. The red points shown on the satellite image refer to the samples collected from the study area.

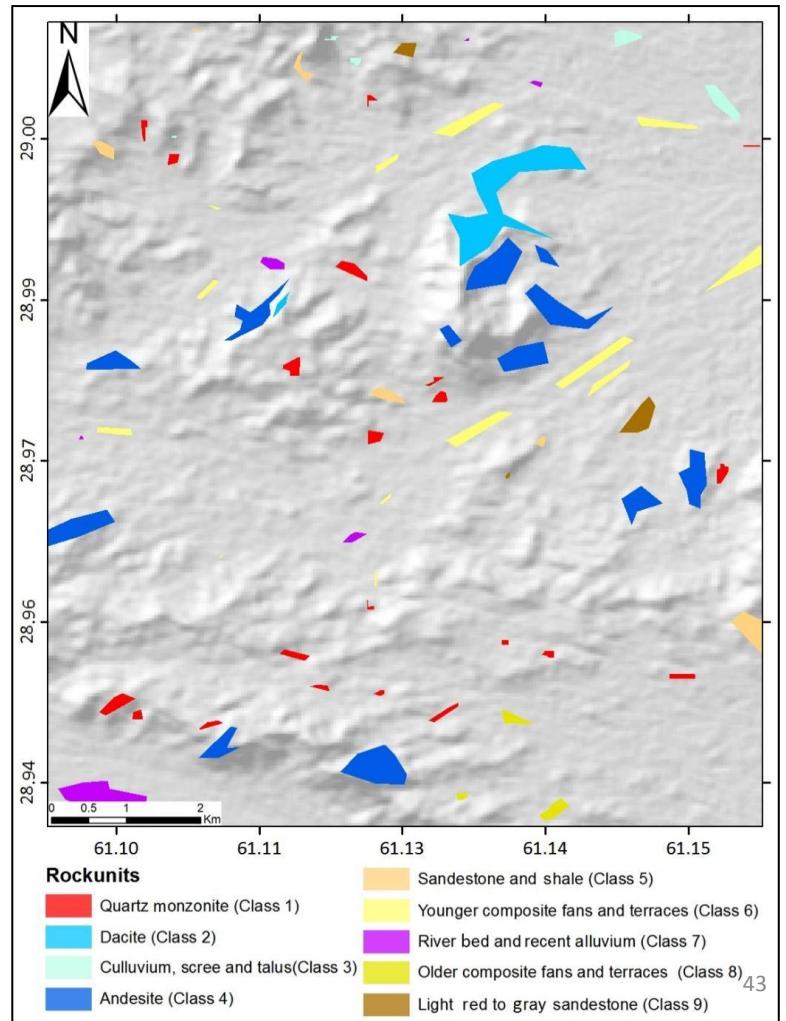




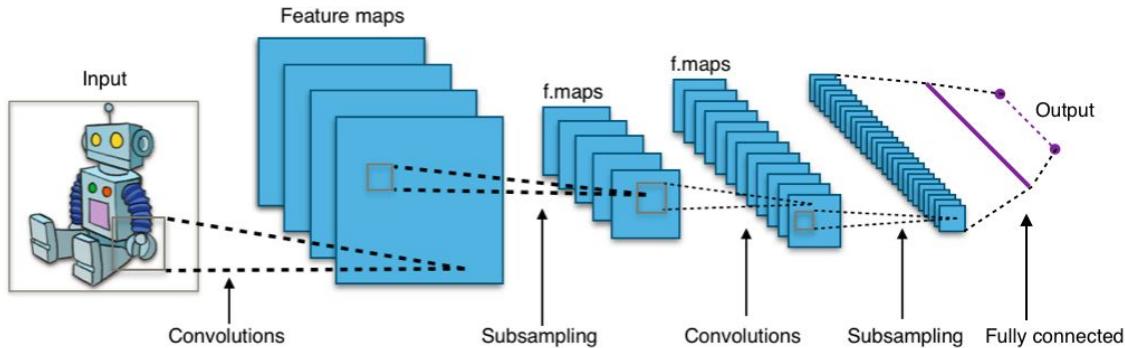
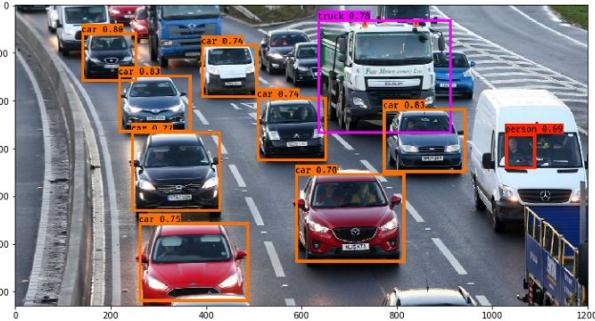
**Figure.** An overview of exposed quartz monzonite rock units within the study area. The photo has been taken close to the sampling location of sample D.



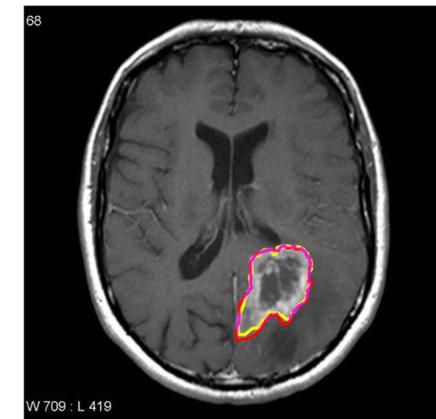
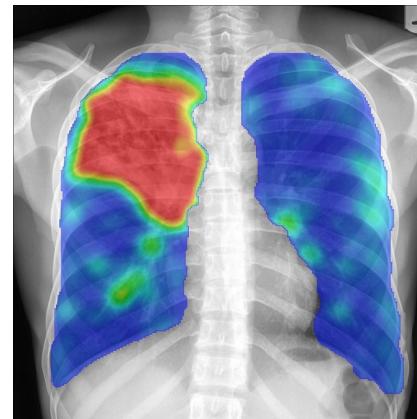
**Figure.** Sample C taken from the quartz monzonite unit and the microscopic sections. In the upper section, plagioclase, hornblende, and epidote phenocrysts are obvious and in the lower section, quartz phenocrysts accompanied by the radial secondary growth of quartz can be seen.



# Deep learning via Convolutional Neural Networks



Source: [https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](https://en.wikipedia.org/wiki/Convolutional_neural_network)



# Multispectral and hyperspectral data

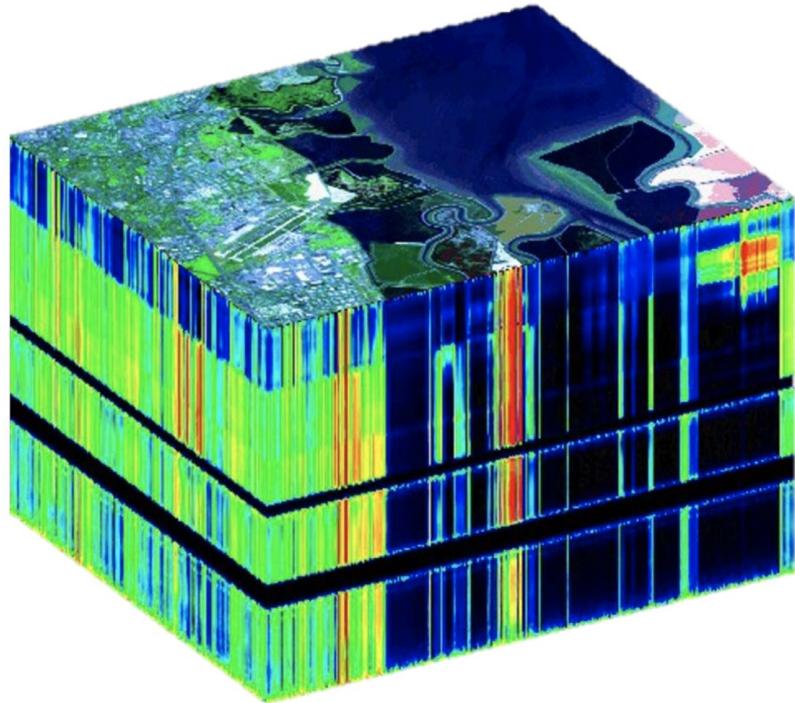
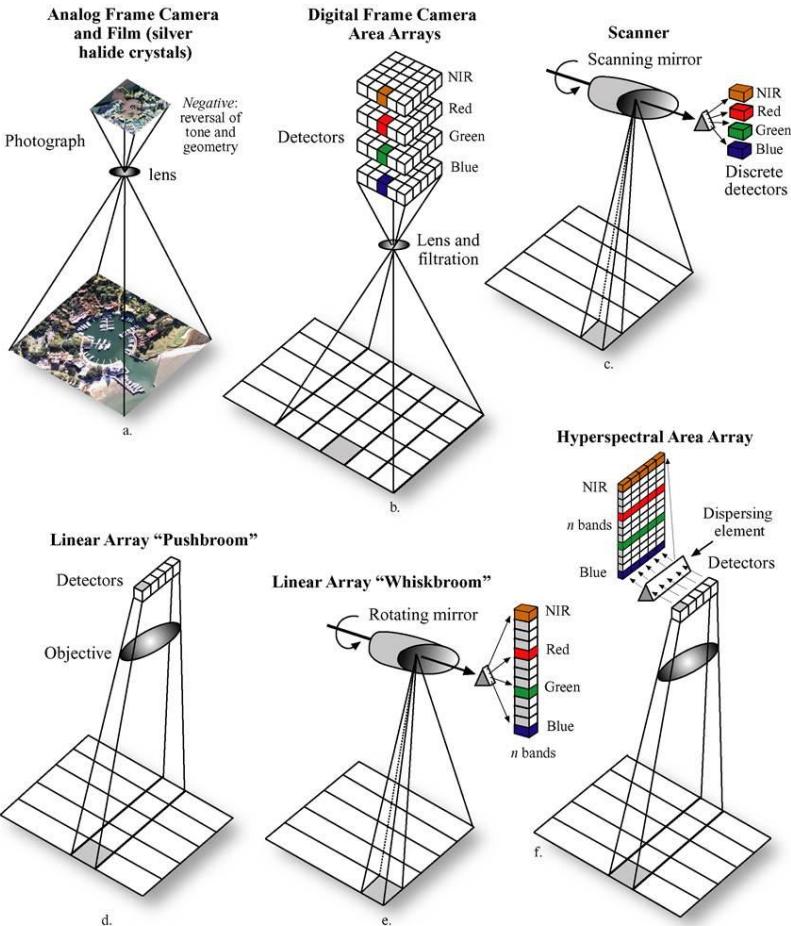
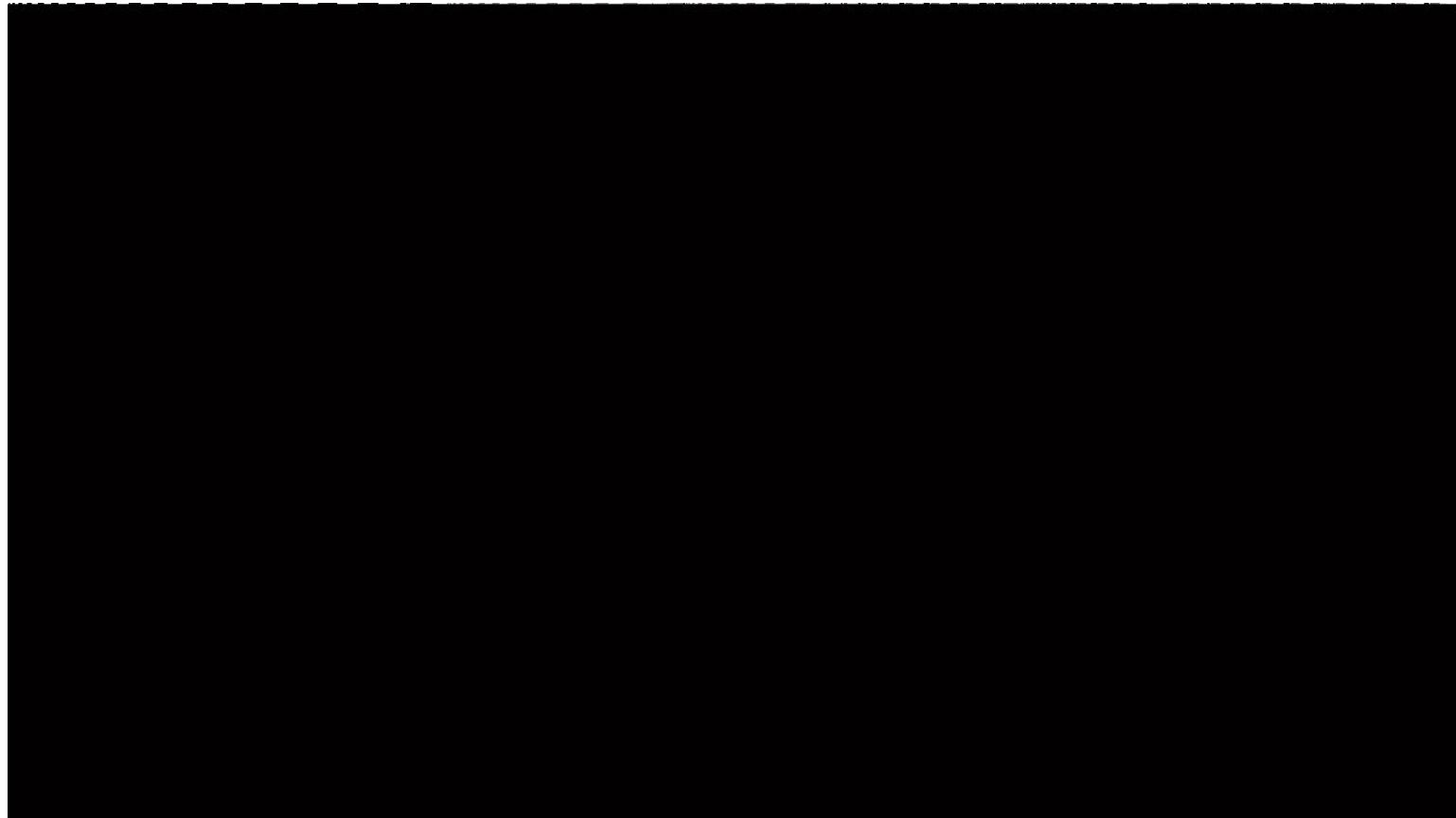


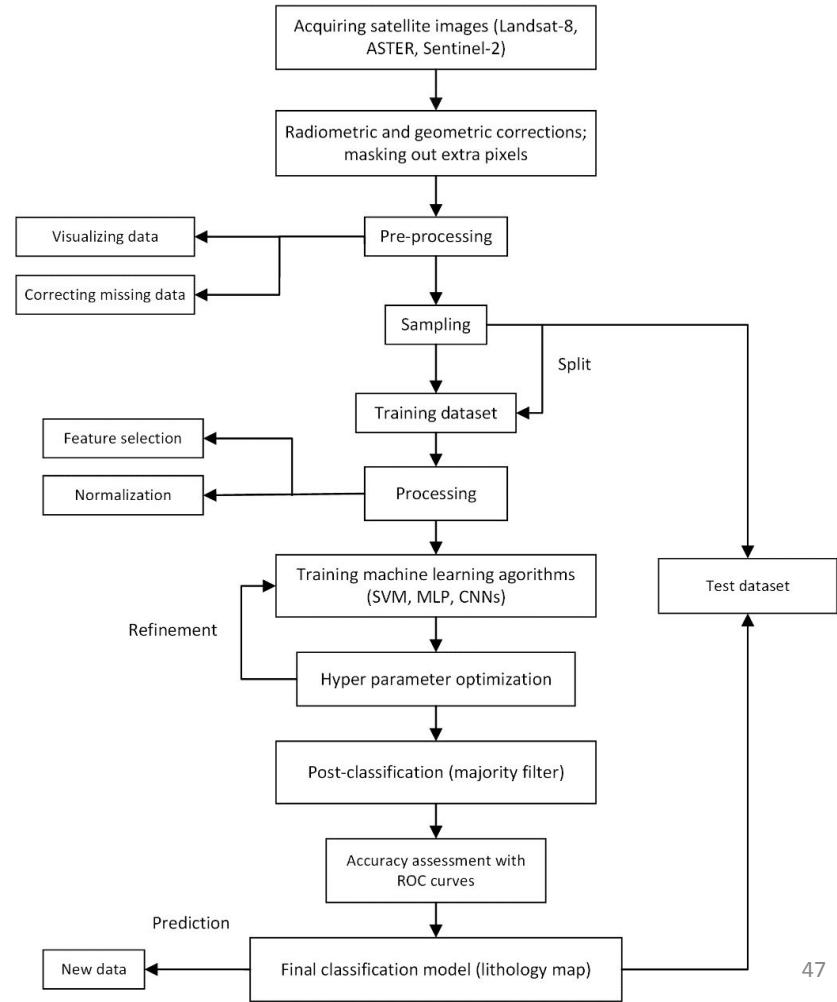
Image Credit: [NASA JPL](#)

## Remote Sensing Systems Used to Collect Aerial Photography, Multispectral and Hyperspectral Imagery

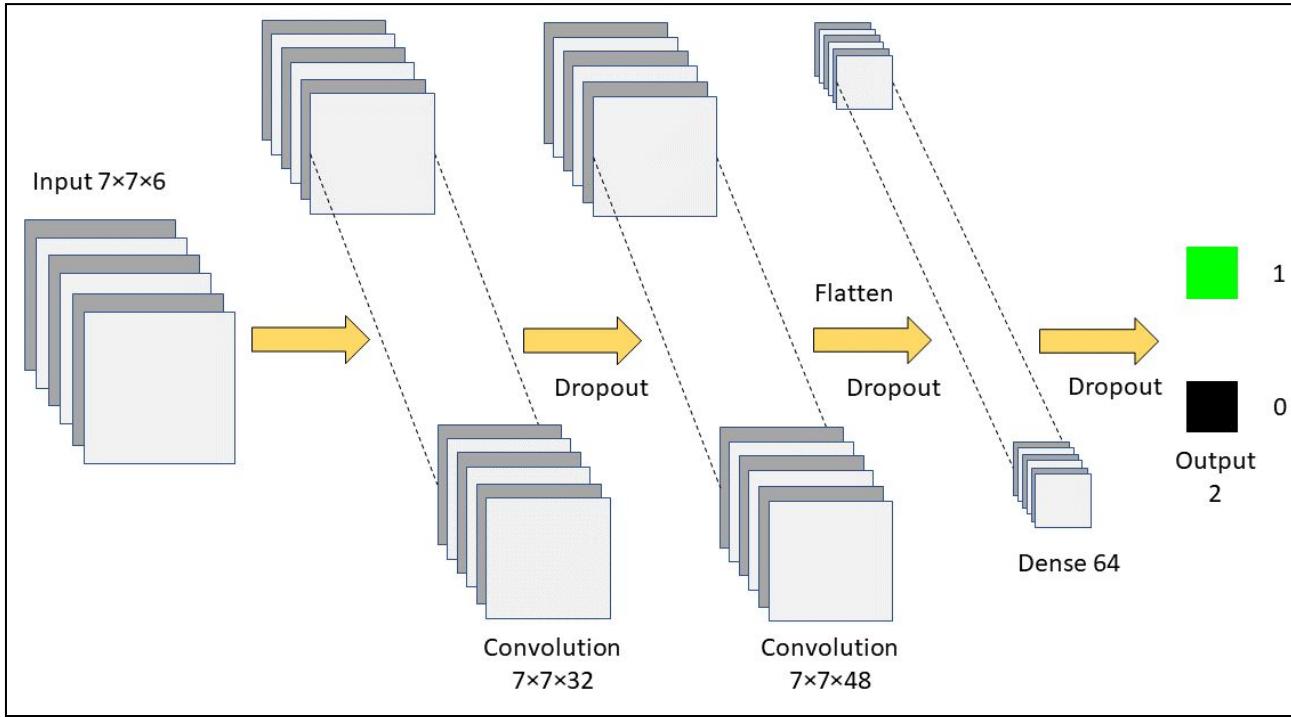




**Figure.** An illustration of generating training chips for a CNN model using the selected bands of Landsat 8 with a  $5 \times 5$  kernel and a  $3 \times 3$  stride.

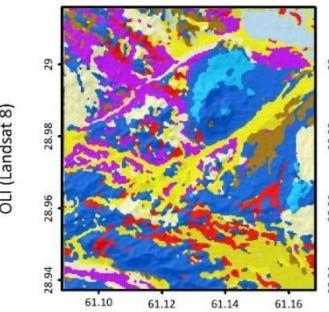


**Figure.** Proposed framework for applying SVM, MLP, and CNN on remote sensing data and mapping lithological units.

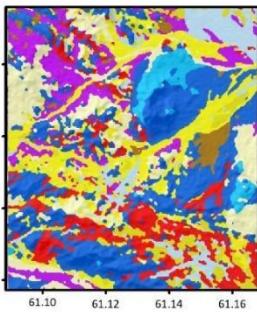


**Figure.** CNN model architecture adopted for classifying the selected bands of Landsat 8 data.

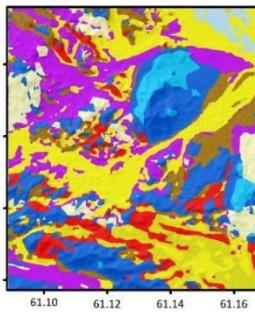
Support Vector Machine



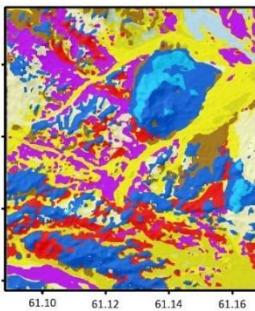
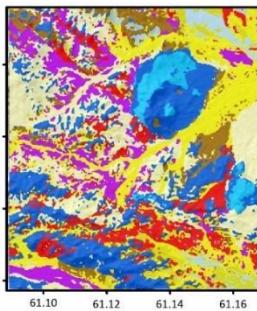
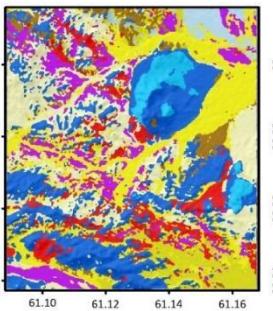
Multilayer Perceptron



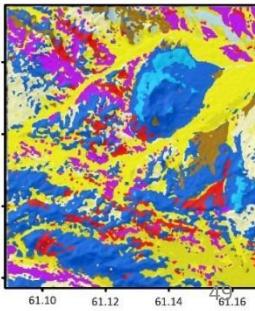
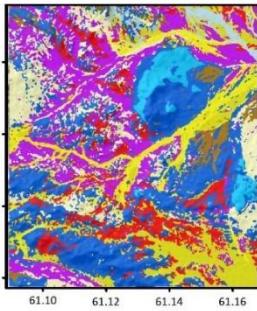
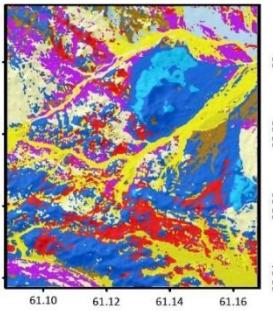
Convolutional Neural Networks



ASTER



Sentinel 2A



**Figure.** Classified maps obtained by applying support vector machine, multilayer perceptron, and convolutional neural networks on OLI (Landsat-8), ASTER, and Sentinel-2 satellite images. The legend is the same as the map provided in Figure 2.

		OLI (Landsat 8)			ASTER			Sentinel-2		
Lithology type	Class number	SVM	MLP	CNN	SVM	MLP	CNN	SVM	MLP	CNN
Quartz Monzonite	1	0.98	0.98	1	0.97	0.99	1	0.88	0.98	0.99
Dacite	2	0.99	0.99	0.99	0.88	0.99	1	0.89	0.99	0.99
Colluvium Scree and Talus	3	0.97	0.97	0.99	1.00	0.99	0.99	0.99	0.99	0.99
Andesite	4	0.91	0.91	0.99	0.90	0.95	0.99	0.80	0.91	0.97
Sandstone and Shale	5	0.97	0.97	0.99	0.89	0.97	1	0.94	0.96	0.98
Younger Composite Fans and Terraces	6	0.92	0.92	0.99	0.94	0.94	0.99	0.95	0.94	0.99
River Bed and Recent Alluvium	7	0.95	0.95	0.99	0.96	0.99	1	0.92	0.92	0.99
Older Composite Fans and Terraces	8	0.90	0.90	1	0.95	0.90	0.99	0.95	0.91	0.99
Light Red to Gray Sandstone	9	0.95	0.95	0.99	0.97	0.98	1	0.96	0.94	0.99

**Table..** The accuracy of each class (lithological unit) obtained by applying SVM, MLP, and CNN on OLI (Landsat-8), ASTER, and Sentinel-2 satellite images.

## Discussion

We showed the efficiency of applying machine and deep learning techniques on satellite data.

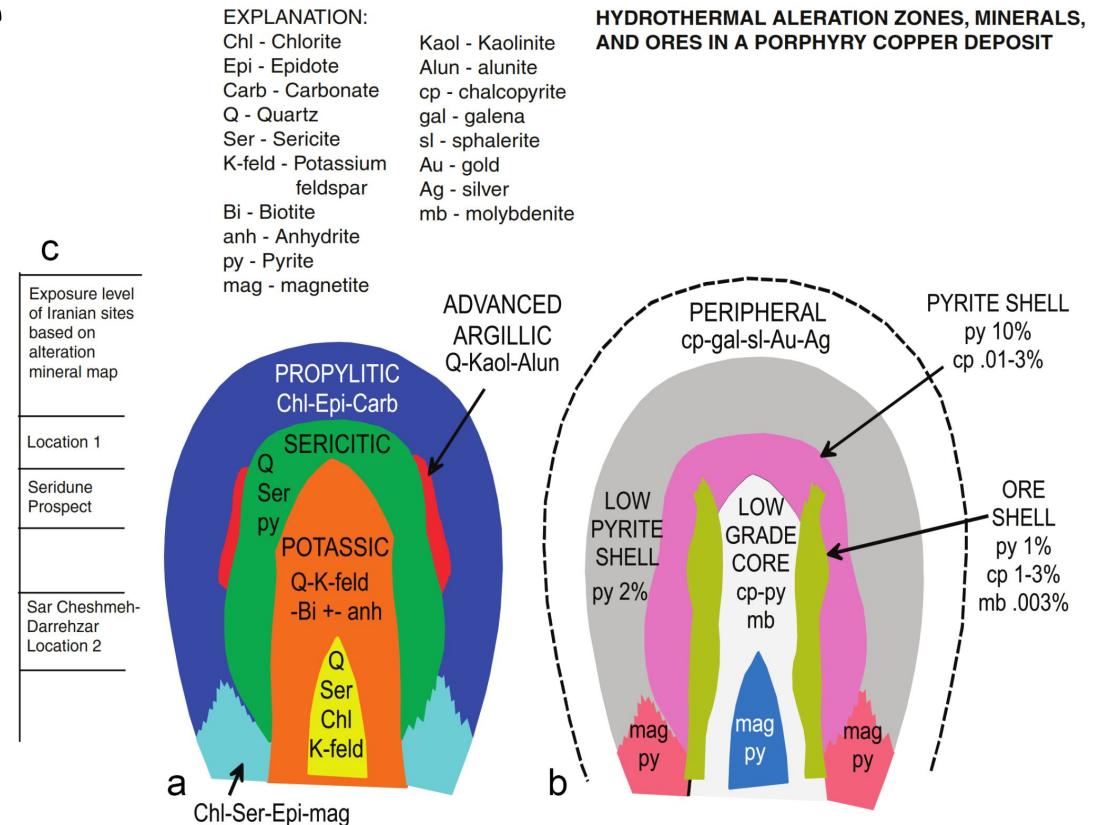
We observed that the combination of CNN and ASTER data provides the most accurate lithological map of the study area based on the ROC curves. Such maps can be considered as base maps for further geological fieldworks and a reliable factor aiding in making decision for mineral exploration operations.

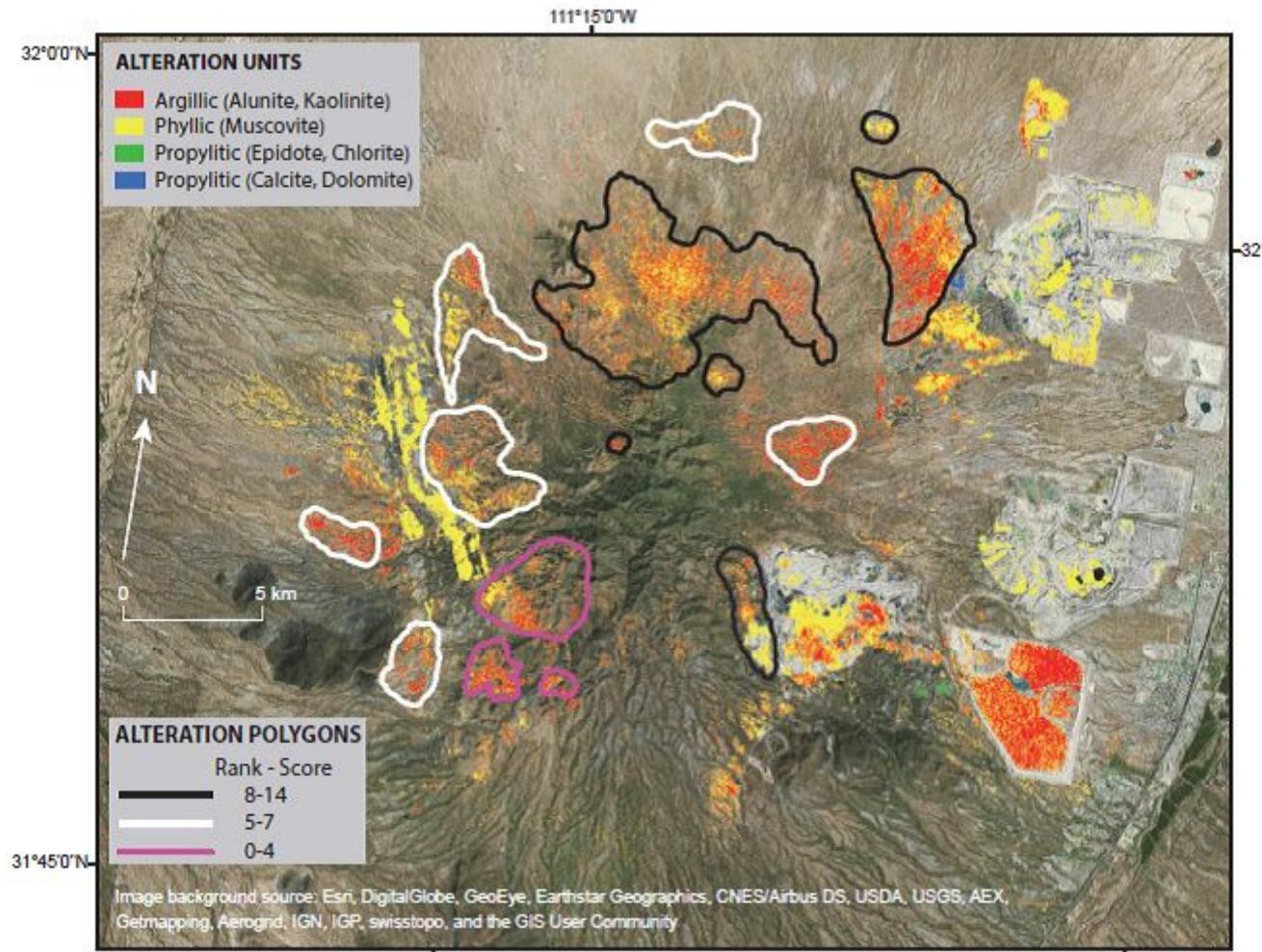
The results show that CNNs and conventional machine learning methods are effective using the respective satellite data sources in generating an accurate lithological map of the study area. However, the combination of CNNs and ASTER satellite data provides the best performance and the highest accuracy and adaptability with field observations and laboratory analysis results.

Our framework presented in a Jupyter notebook is an open-source community tool for mapping lithological units using any multi- or hyper-spectral data. This notebook can significantly enhance the ability of exploration geologists to map lithological units.

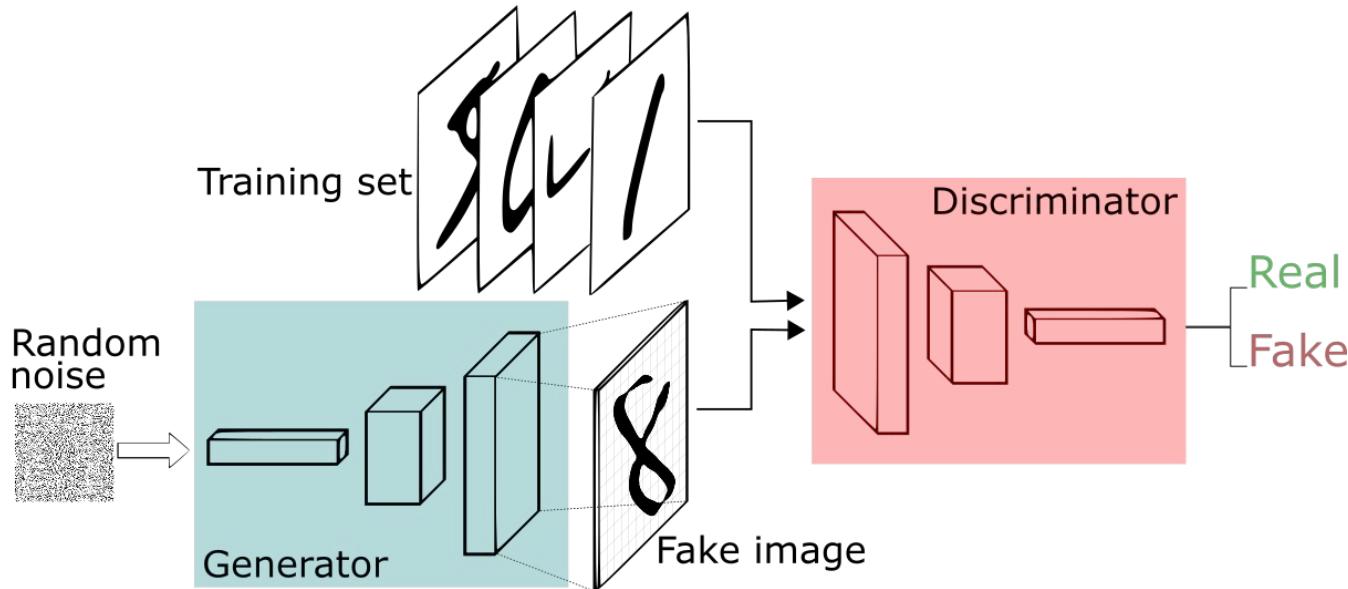
# Convolutional Neural Networks and Remote Sensing for Alteration Zones

Dhiraj Pimparkar and Dakshi Goel (Indian Institute of Technology - Jammu)





# Data Augmentation: Generative Adversarial Networks (GANs)



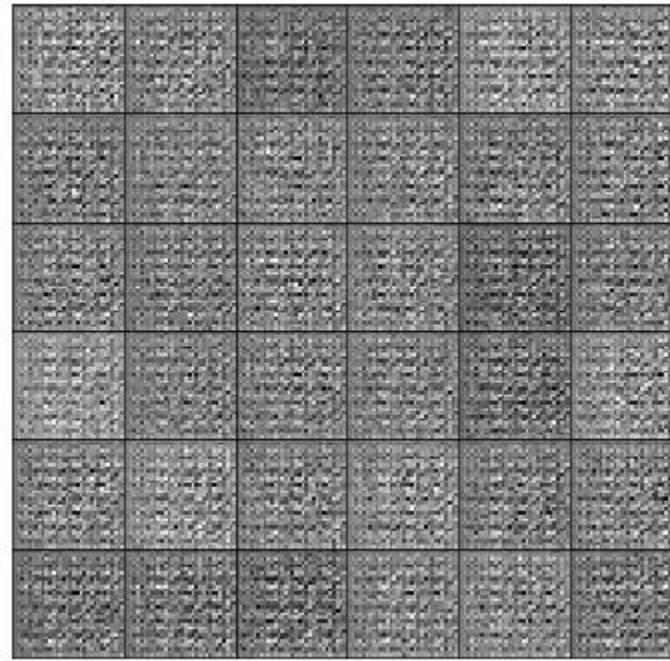
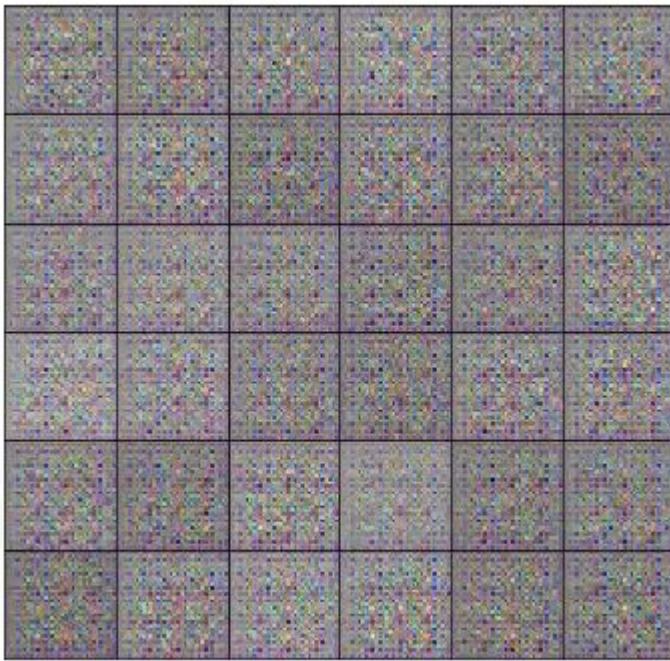
Source:

<https://towardsdatascience.com/generative-adversarial-networks-gans-explained-101-102-103-104-105-106-107-108-109-10a-10b-10c-10d-10e-10f-10g-10h-10i-10j-10k-10l-10m-10n-10o-10p-10q-10r-10s-10t-10u-10v-10w-10x-10y-10z>

In the GAN framework:

- The generator maximizes the probability of making the discriminator mistakes its inputs as real.
- The discriminator guiding the generator to produce more realistic images.

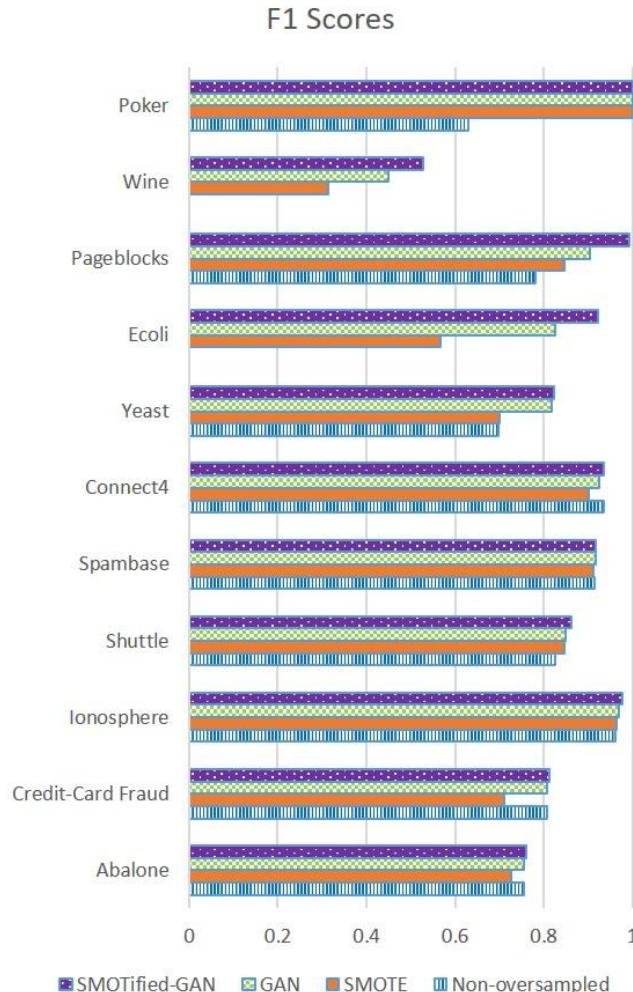
In the perfect equilibrium, the generator would capture the general training data distribution. As a result, the discriminator is always unsure of whether its inputs are real or not.



*"We used a 4 layer convolution network for (both discriminator and generator) with batch normalization to teach a model to generate SVHNs and MNIST images. Above, are the SVHN's (left) and MNIST (right) generator samples during training."* Source:  
<https://sthalles.github.io/intro-to-gans/>

## Pattern classification (non-image datasets)

Can be extended to  
non-Image data?



Anuraganand Sharma, Prabhat Kumar Singh, Rohitash Chandra,  
“SMOTified-GAN for class imbalanced pattern classification problems”.  
<https://arxiv.org/submit/3871315/view>

The road ahead ....



# References

1. Chandra, R., Cripps, S., Butterworth, N., & Muller, R. D. (2021). Precipitation reconstruction from climate-sensitive lithologies using Bayesian machine learning. *Environmental Modelling & Software*, 139, 105002.
2. Shirmard, H., Farahbakhsh, E., Müller, R. D., & Chandra, R. (2022). A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, 268, 112750.
3. H. Shirmard, E. Farahbakhsh, E. Heidari, A. B. Pour, B. Pradhan, R. D. Müller and R. Chandra. A comparative study of convolutional neural networks and conventional machine learning models for lithological mapping using remote sensing data. *Remote Sensing* (to appear in January 2022)
4. J Diaz-Rodriguez, RD Muller, R Chandra. Predicting the emplacement of Cordilleran porphyry copper systems using a spatio-temporal machine learning model. *Ore Geology Reviews* 137, 2021: 104300
5. Chandra, R., Jain, M., Maharana, M., & Krivitsky, P. N. (2021). Revisiting Bayesian Autoencoders with MCMC. *arXiv preprint arXiv:2104.05915*. <https://arxiv.org/abs/2104.05915>
6. A Sharma, Prabhat Kumar Singh, R Chandra, "SMOTified-GAN for class imbalanced pattern classification problems". <https://arxiv.org/submit/3871315/view>
7. S. Chandra, M Lindsay, S Clark, and R. Chandra, Drill-core analysis for mineral composition with an unsupervised machine learning framework, *Ore Geology Reviews*, 2022 (In review)
8. Chandra R; Azam D; Kapoor A; Dietmar Müller R, 2020, 'Surrogate-assisted Bayesian inversion for landscape and basin evolution models', *Geoscientific Model Development*, vol. 13, pp. 2959 - 2979
9. Pall J; Chandra R; Azam D; Salles T; Webster JM; Scalzo R; Cripps S, 2020, 'Bayesreef: A Bayesian inference framework for modelling reef growth in response to environmental change and biological dynamics', *Environmental Modelling and Software*, vol. 125, pp. 104610 - 104610
10. Chandra R; Müller RD; Azam D; Deo R; Butterworth N; Salles T; Cripps S, 2019, 'Multicore Parallel Tempering Bayeslands for Basin and Landscape Evolution', *Geochemistry, Geophysics, Geosystems*, vol. 20, pp. 5082 - 5104

## Open Source Software

All papers provide open source software and data via Github repository

Many thanks to everyone for attending and special thanks to everyone behind the scenes.