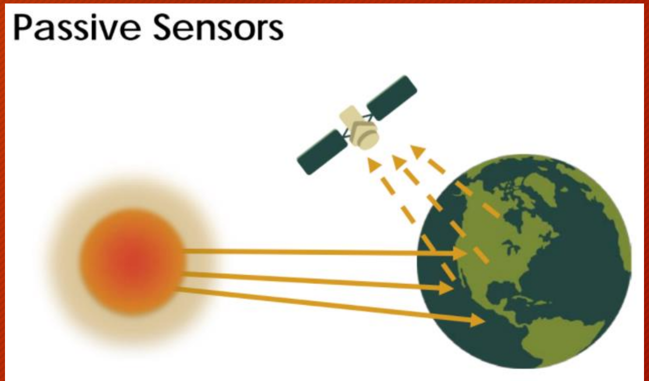
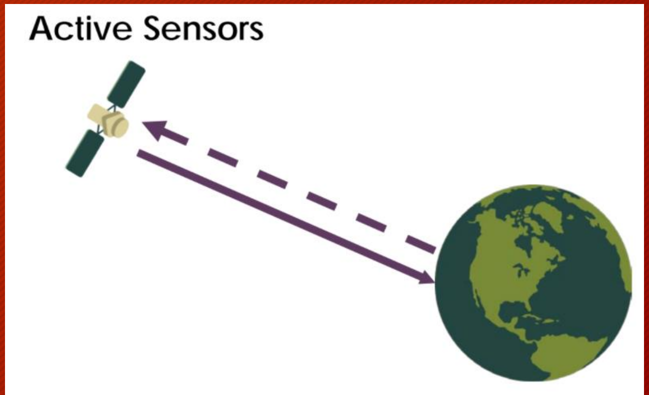
Tech Brief:
Optimizing Machine Learning Techniques for Earth Observation Data

Paul Struhsaker
Arrasar Partners
www.struhsaker.com/arrasar-partners

Earth Observation (EO) Introduction

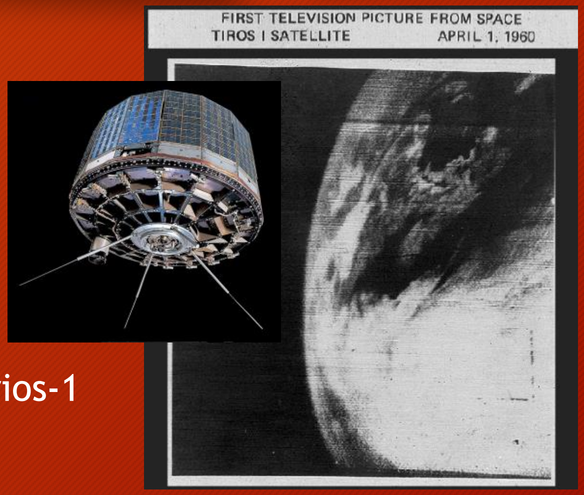
- Earth Observation is conducted using Active and/or Passive Sensors
  - **Passive Sensors:** Based on solar reflection in visible, infrared, thermal infrared, and uwave bands of EM spectrum.
    - Cloud coverage and night side of the earth can limit operation
  - **Active Sensors:** Based on satellite emitting and receiving reflected energy
    - Synthetic Aperture Radar (SAR) – for 3D imaging
    - Generally all-weather operation
- Key Systems Parameters for EO satellite
  - **Spatial Resolution:** size of object/features being measured
  - **Spectral Extent:** breadth of electromagnetic spectrum sensed
  - **Radiometric Resolution**: number of digital levels expressed in the data collected
  - **Temporal Resolution**: intervals between imagery acquisition



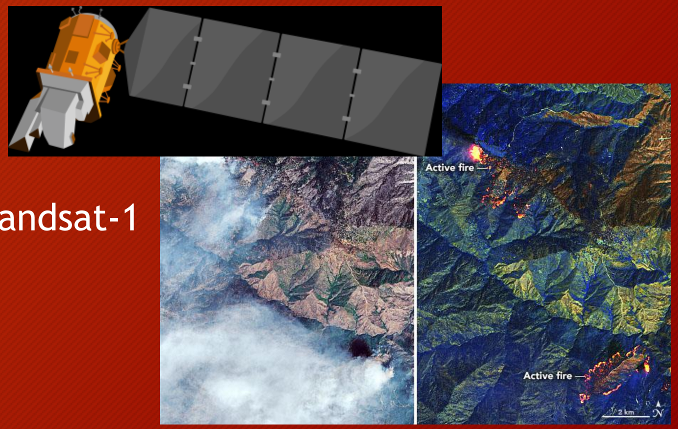


Earth Observation a brief history

- 1960 Tiros-1 weather satellite
  - 2 visible spectrum cameras – 10 satellites launched
- 1964 – Nimbus-1 to 1978 Nimbus 7
  - introduced multi-spectral scan (MSS) w/uwave for water vapor detection
- 1974 – Landsat 1 to 2013 Landsat 8
  - MSS progressed from 3 bands to 11 bands
- 1974 - GEOS series - 1st Geostationary orbit EO satellites
  - GEOS-17 launched 2018 –<u>program on going</u>
- 1980's Active Sensing imaging … SAR introduced
- 1986 SPOT program – commercial earth observations
- 1990 -2010 sensor types expand
  - 10 countries/regions (EU) operate EO/meteorological satellites
- 2010's – Nano/cube Sats – miniaturization w/full capabilities



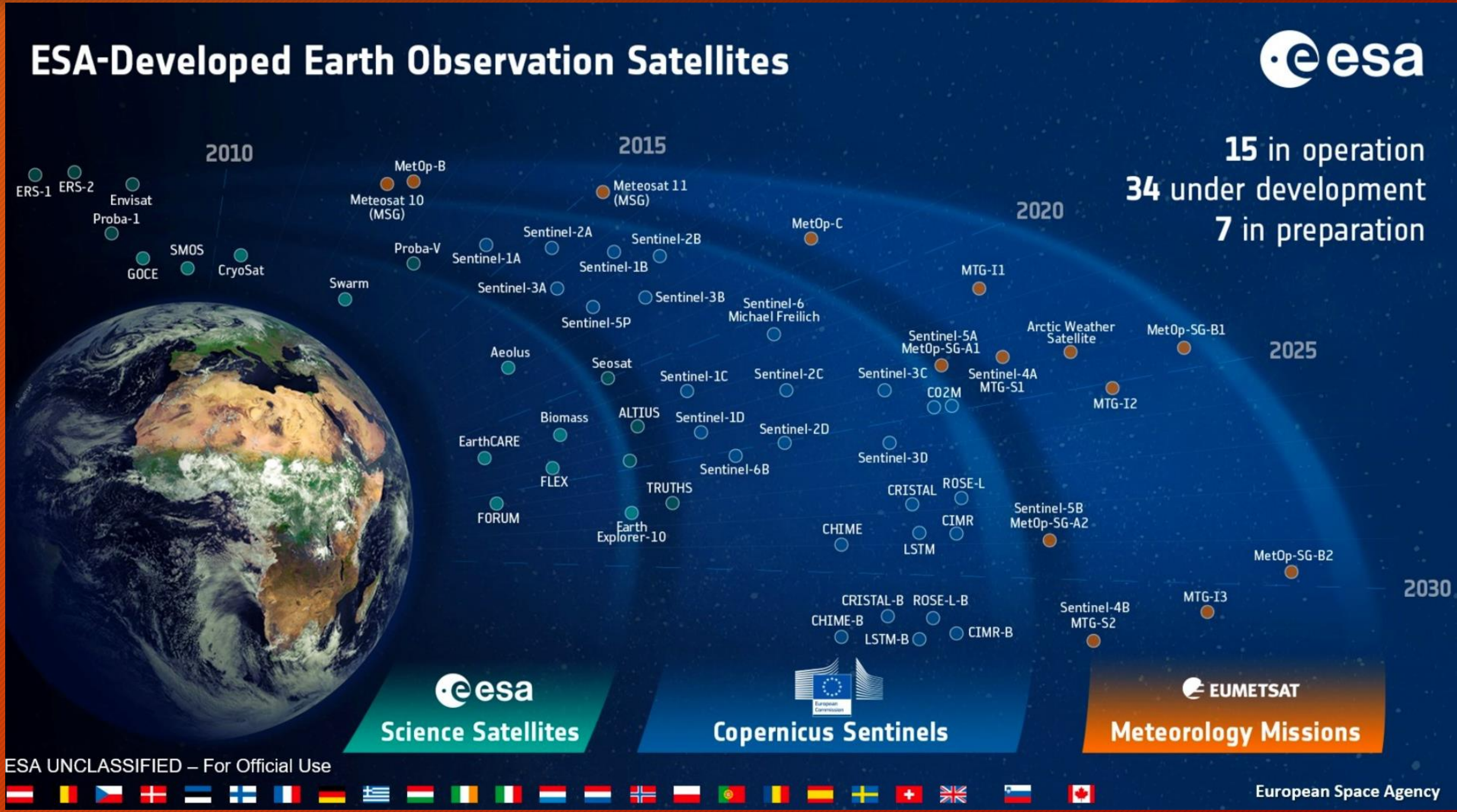
Trios-1



Landsat-1

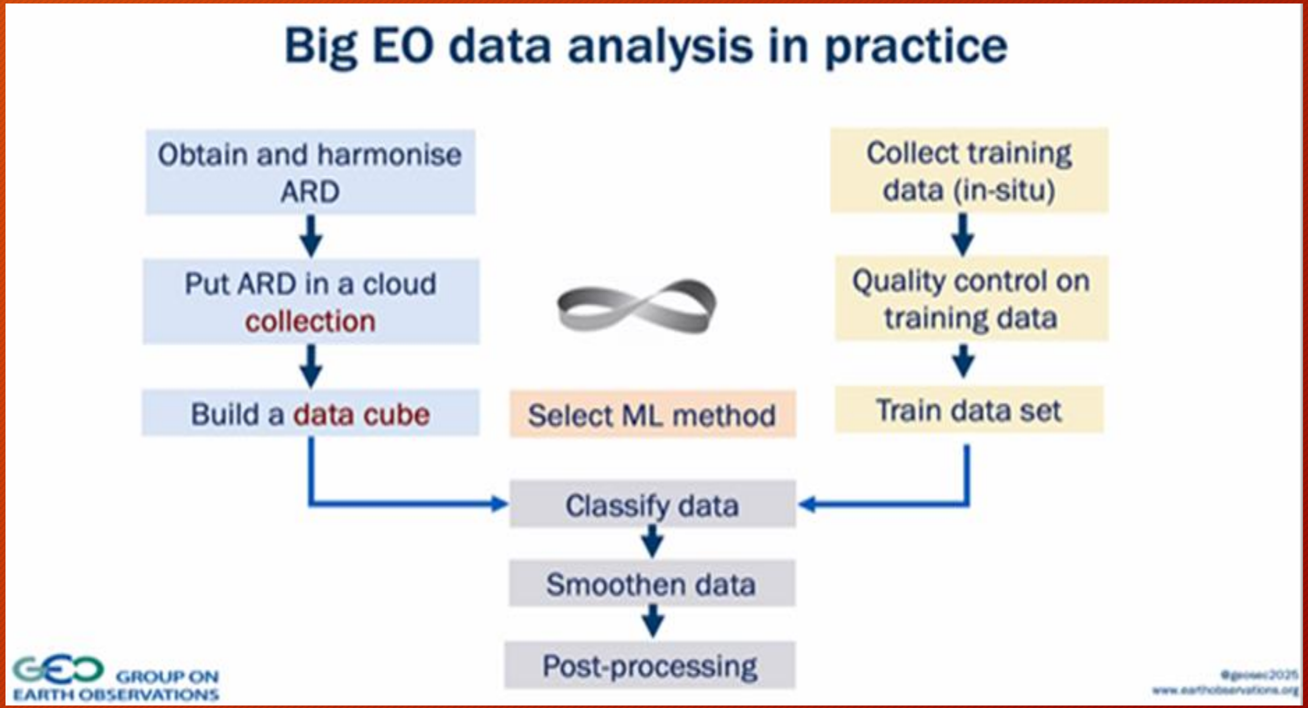
The Earth Observation Data Problem:
The incredible volume of data generated daily

- ESA alone has 15 operational EO platforms with 7 more on the way
- Commercial imaging satellites are collecting over 100 Terabytes a day
- Petabytes of data generated daily - it will soon reach Exabytes
- Large scale Machine Learning are the only practical way to make use of the deluge of data



Machine Learning of Earth Observation Data - Overview

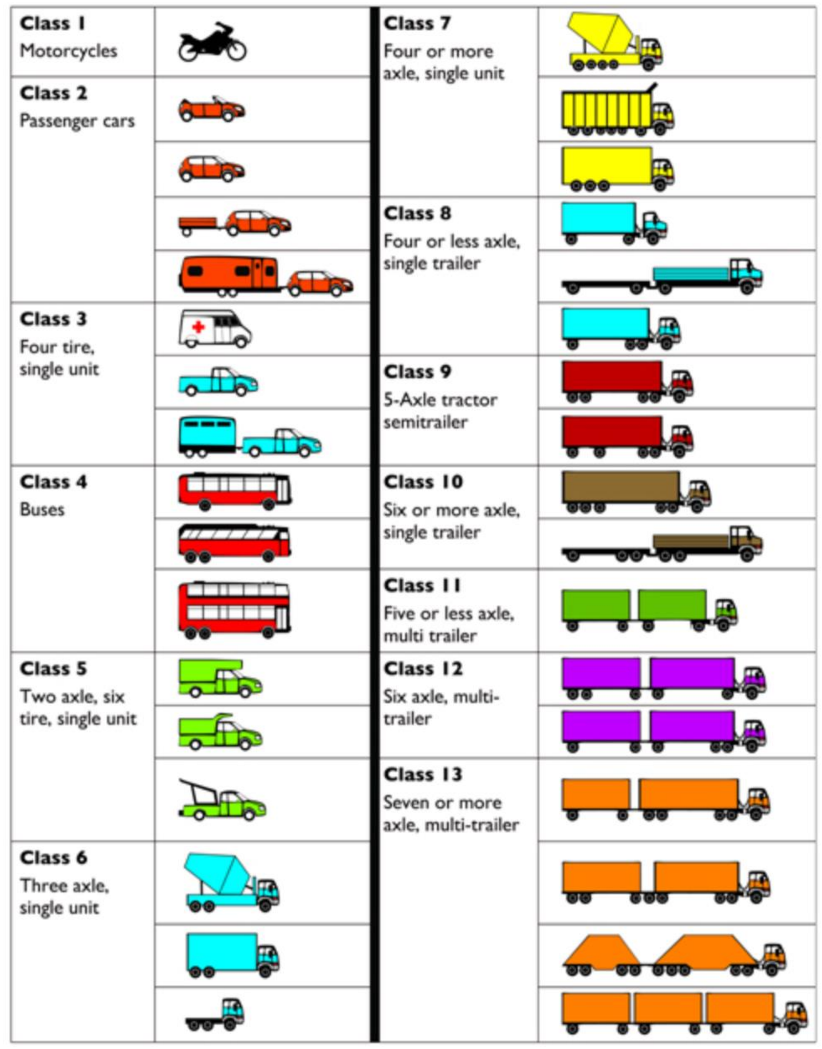
- Machine Learning requires two parallel processes
- First, vast amounts of data is captured from EO satellites
  - EO data is processed to make it application-ready data (**ARD**)
  - ARD is stored in the Cloud and organized into subsets called **Data**
- Secondly, ML models are trained with specific training data to detect desired features or events
  - Quality control and training of ML model is critical
- With datasets organized, the ML model is selected to classify, smoothen, and yield results



Data Labeling: Dealing with Sheer Volume of Data

- The vast amount of satellite EO makes efficient use of ML models and algorithms difficult
  - More than 2 petabytes (PB) of EO data generated daily from form Sentinel 2/3, Landsat 8, etc.
  - Date can be 2D or 3D and contain multiple spectral samples
  - Many ML models operate efficiently on sample models, but fail or perform poorly on live data sets
- Labeling data critical & contains two elements
  - <u>**Localization:**</u> defined area or "box" of an image
  - <u>**Classification:**</u> assign one of a predetermined set of categories
    - Critical to a have an exhaustive detailed set of Labels
    - Example- vehicle classification: car, SUV, Truck, Van, etc.

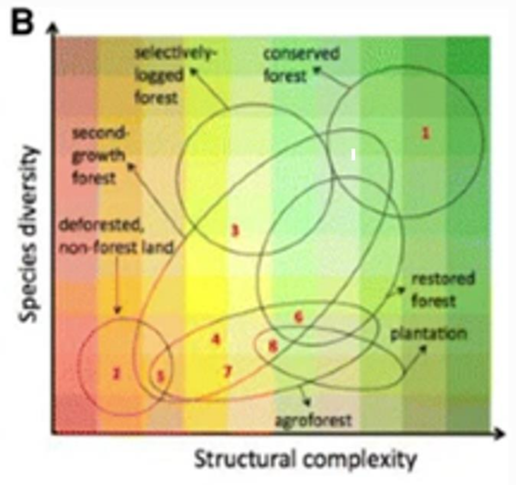
Vehicle Classifications



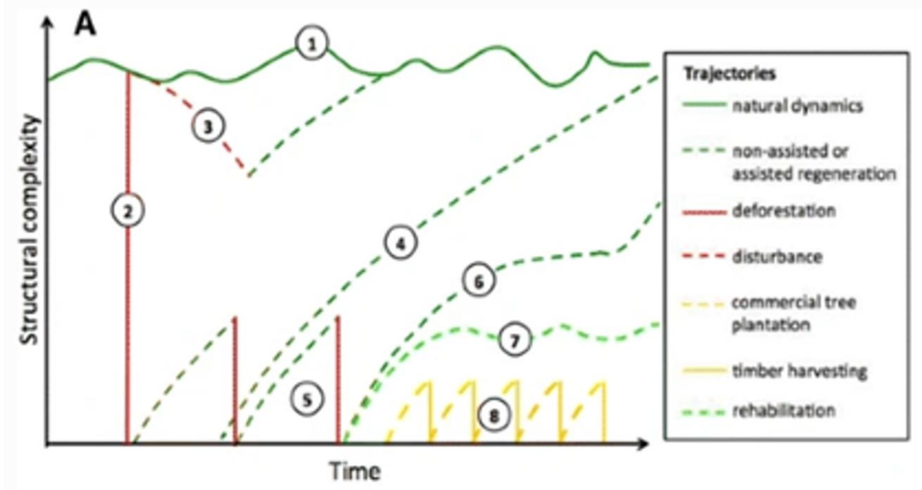
Source: Federal Highway Administration

Time as a critical element in ML analysis

- Many EO ML tasks analyze changes over time as examples:
  - Forest management and forest health (see graphic)
  - Home solar installation growth in USA
- Data must be stored with Geospatial Semantics
  - Time is included with 2D / 3D data
  - Depending on satellite orbit – revisit time is variable
- Unfortunately, many ML classifiers and algorithms perform poorly in a changing environment
  - Spatial analysis vs Time-Series analysis



Forest Structure



Forest Life Cycle

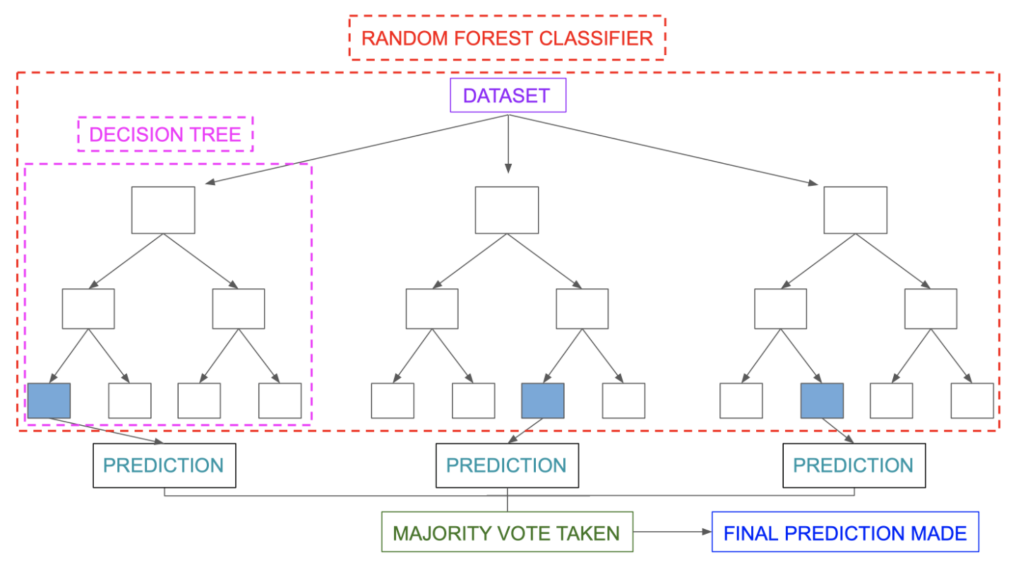
Additional Data Optimization Challenges

- Resolution and data format
  - Satellites have different resolution – 500m to 0.3m
  - Different satellite systems format to different specifications JPEG2000, GeoTIFF -
  - One solution, use **Sentinel Hub provides** a single format for Sentinel-1/2,Landsat 8, and MODIS
- • The Cloud Issue:
  - Satellite images can partly or fully covered by clouds
  - Images with clouds need to be detected and masked from processing
- Accuracy of image georeference points

Resolution of EO satellite system

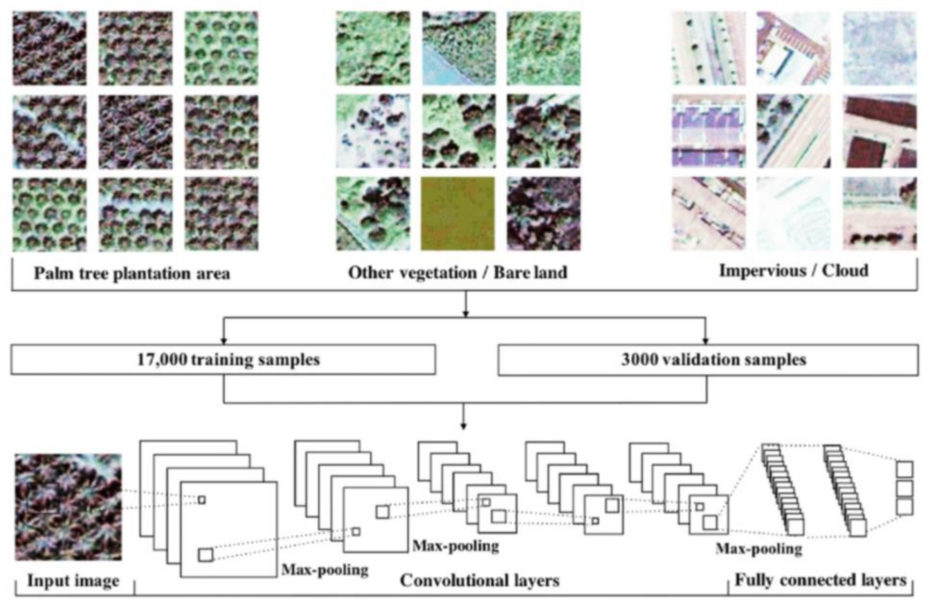
Machine Learning Algorithms

- Machine Learning really is learning by example, so it is critical to provide as complete a training data set as possible
  - Performance is highly dependent on the COMPLETENESS of the training set
- Typically, one of three algorithms used for analysis
  - *Multivariate non-linear non-parametric regression*
    - Decision Tree, Random Forest, etc.
  - *Supervised classification*
    - *Trained Neural Networks, & Support Vector Machines*
  - *Unsupervised classification*
    - Untrained Neural Network & Support Vector Machines

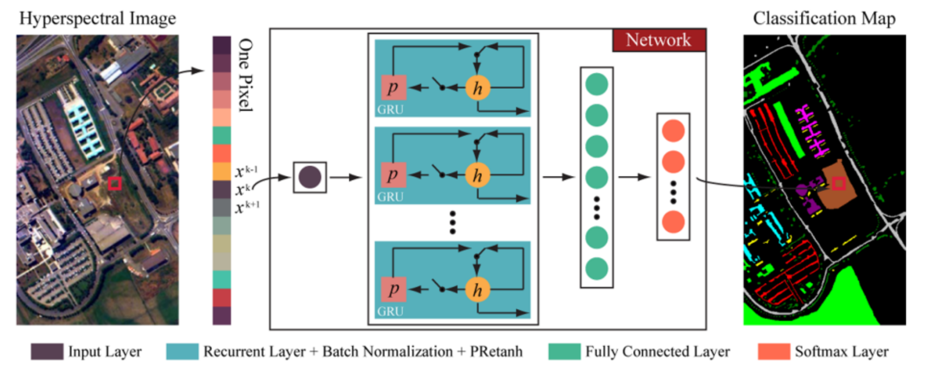


Neural Networks: CNN & RNN

- CNN Conventional Neural Network
  - CCN used extensively in image recognition
  - Network consists of three layers
    - Convolution Layer – preprocess data (~optic neuron)
    - Pooling Layer(option) – clusters data to next layer
    - Receptive Field – reduces inputs to next layer
    - Weights – applied and filter data distinguish the image feature
- RNN Recurrent Neural Network
  - RNN Processes data serially
  - Ideal for used text/speech recognition
  - RNN Hyperspectral Image Classification: analyze hyperspectral pixels as sequential data and then determine information categories via network reasoning



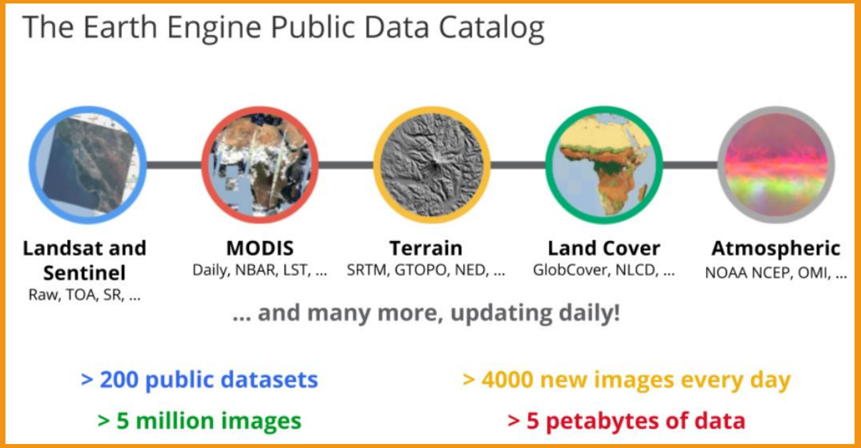
CNN-Conventional Neural Network

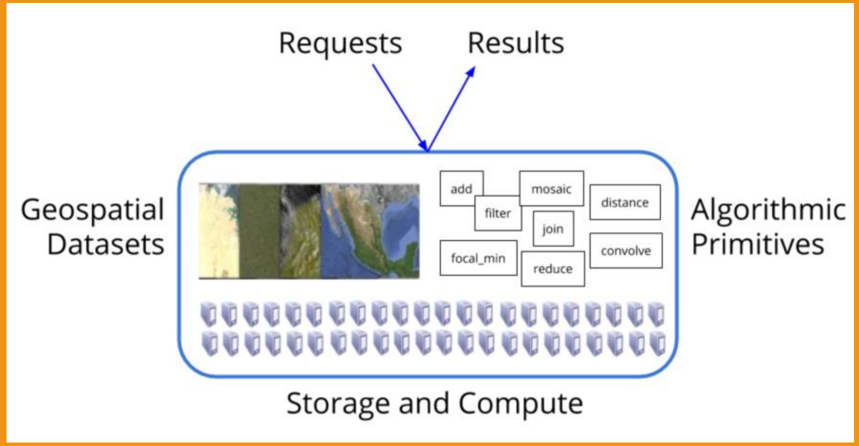


RNN-Recurrent Neural Network (hybrid)

Earth Observation Tool Kits

- Google Earth Engine (GEE)
  - GEE provides terabytes of satellite imagery sources, data processing tools and ML algorithms
  - Custom ML models can be developed and run in the GEE browser and Python is supported
- EO-learn
  - Python package that links closely with the data science and machine learning python ecosystem to the remote sensing/earth observation community
- Radiant MLHub
  - Offers ready to use training datasets in SpatioTemporal Asset Catalog (STAC) format developed by NASA
  - training datasets cover different machine learning applications including image classification, segmentation and object detection





Thank You

Questions/Comments Contact: paul@struhsaker.com
Visit the Arrasar Web Page: www.struhsaker.com/arrasar-partners
PDF download available at: www.struhsaker.com/dystopic-science-stories