

Applying Artificial Intelligence Techniques to Earth Observation Data Quality Control Activities

Andrea Melchiorre, Ruben Columina,
Fay Done, and Kevin Halsall





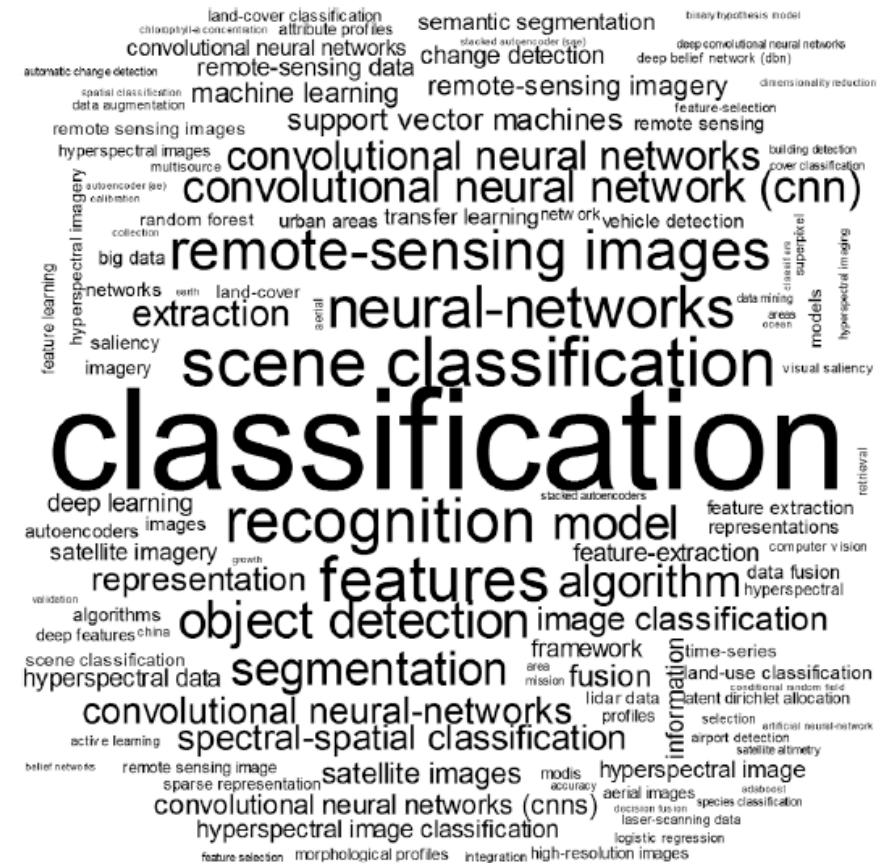
Summary

- > Telespazio VEGA UK have established a process to create AI models to detect anomalies in optical EO datasets
- > The AI models have the potential to support EO QC engineers against a growing number of EO missions and data product volumes
- > A tool has been developed to capture the output of an EO QC engineer's analysis to support AI model development
- > An AI model has been developed to support the QC of ESA Landsat products for the IDEAS+/QA4EO Service



EO Data Quality Control

- Telespazio VEGA UK are prime contractors for ESA's IDEAS+ service – responsible for performing operational and offline QC analyses on many ESA and Third Party EO data
- QC activities are a mixture of automated checks (applied to the whole dataset) coupled with a more detailed human observation of a smaller subset of the data
- Year on year, data volumes are increasing – very hard for available resources to increase at the same pace
- The use of AI/Machine learning technics for classification is increasing exponentially
- QCOLT (Quality Control Optical Learning Tool) project was conceived to determine feasibility of using AI/Machine Learning techniques to support the QC of EO data
- Test application for the QC activity associated with the bulk re-processing over 600,000 ESA Landsat MSS products
- Overall objective was to determine whether the types of detailed QC analysis done by operators to a small data subset could be extended to entire datasets

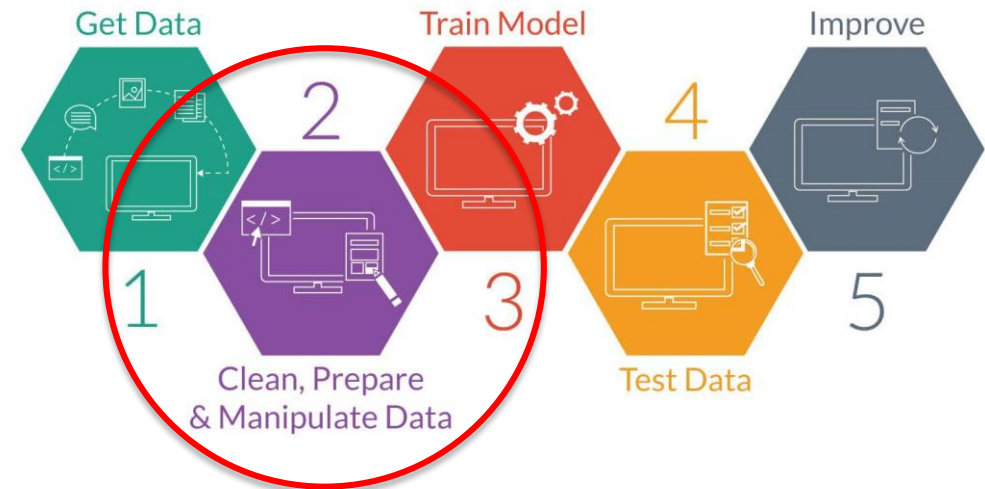


Lei Ma et al. 2019



Machine Learning

- At a high level – development of an Machine Learning model has 5 steps
- **Data Preparation** step is the most effort intensive
- Data Preparation involves:
 - Performing a QC assessment on data products within an archive
 - Tagging those products exhibiting specific anomalies
 - Creating two training datasets consisting of:
 - Nominal data
 - Data containing the specific anomaly being investigated
- Most of this type of assessment is performed by a QC engineer as part of their normal activities



BUT assessments aren't typically captured in a way that AI experts can use to properly prepare the data

- Project developed the **QCOLT software** to support both QC engineers and AI experts:
 - QC engineer duty is to perform large elements of the Data Preparation step for the AI experts
 - The Machine learning output reduces the resources needed to complete the QC over the full dataset



QCOLT Software

- Large **backend** database containing information and image data relating to all of the Landsat products under investigation
 - Product data (images)
 - Product Metadata
 - Product statistics
 - Quality reports from automated analyses outputs

Version: dev-pre-production Q-COLT [Logout](#)

Search LANDSAT Products e.g. "LS1 MSS 1975","LS2 ETM 2000", "LS2 ETM 2000 2003" PRODUCTS PER PAGE 10

[Shuffle for Inspection](#) Total Products Selected: 1401 [Amalfi Filter](#) [Anomaly Filter](#) [Probability Filter](#) [Other Filters](#)

Product Name	Satellite	Instrument	Absolute Orbit	IGS	Product Type	Date	Path	Row	AMALFI Status	Actions
LS01_MSS_19760907T094129_19760907T094158	LS01	MSS	021028	RMTI	GTC	1976-09-07	0218	0022	Passed	Inspection Show Details
LS01_MSS_19760911T100405_19760911T100434	LS01	MSS	021084	RMTI	GTC	1976-09-11	0222	0022	Passed	Inspection Show Details
LS01_MSS_19751105T100023_19751105T100052	LS01	MSS	016747	RMTI	GEO	1975-11-05	0217	0029	Passed	Inspection Show Details
LS01_MSS_19760523T080936_19760523T081006	LS01	MSS	019535	RMTI	GEO	1976-05-23	0201	0015	Passed	Inspection Show Details
LS01_MSS_19760523T080956_19760523T081026	LS01	MSS	019535	RMTI	GEO	1976-05-23	0201	0016	Passed	Inspection Show Details
LS01_MSS_19760523T081021_19760523T081052	LS01	MSS	019535	RMTI	GEO	1976-05-23	0201	0017	Passed	Inspection Show Details
LS01_MSS_19760523T081046_19760523T081116	LS01	MSS	019535	RMTI	GEO	1976-05-23	0201	0018	Passed	Inspection Show Details
LS01_MSS_19760523T081111_19760523T081141	LS01	MSS	019535	RMTI	GEO	1976-05-23	0201	0019	Passed	Inspection Show Details
LS01_MSS_19760523T081136_19760523T081205	LS01	MSS	019535	RMTI	GEO	1976-05-23	0201	0020	Passed	Inspection Show Details
LS01_MSS_19760523T081200_19760523T081230	LS01	MSS	019535	RMTI	GEO	1976-05-23	0201	0021	Passed	Inspection Show Details

« < 1 2 3 4 ... > »



QCOLT Software

- Customised **GUI** permitting:
 - visual inspection of the data
 - inspection of metadata
 - anomaly assignment
- During inspection the training dataset is built
- Once AI models implemented, results imported back into QCOLT

The screenshot shows the QCOLT software interface. At the top, there's a header with 'Version: dev-pre-production', the 'TELESPAZIO a LEONARDO and THALES company' logo, and a globe icon with '-COLT'. Navigation buttons include 'Administration', 'Short cuts', and 'Logout qcolt'. Below the header, there are tabs for 'Images', 'Inspection', 'Resources', and 'Region of interest'. The 'Resources' tab is active, showing a satellite image of a coastal area. To the right of the image are buttons: 'Clear', 'Reset', 'Cycle QL', 'Use Hi Res image', and 'Export'. On the far right, there's a control panel with 'Mark as Not Inspected', '1 of 613704', and 'Selection' buttons. Below that is a table for 'Anomalies Assigned':

Anomalies Assigned	
<input type="radio"/> miRegularNoise	0.00%
<input type="radio"/> OK_Candidate	
<input type="radio"/> Double_check	
<input type="radio"/> miScanStart	100.00%
<input type="radio"/> unfigd_dropped	
<input type="radio"/> Ls03_mirrored	
<input type="radio"/> SV issue	
<input type="radio"/> CloudIssue	

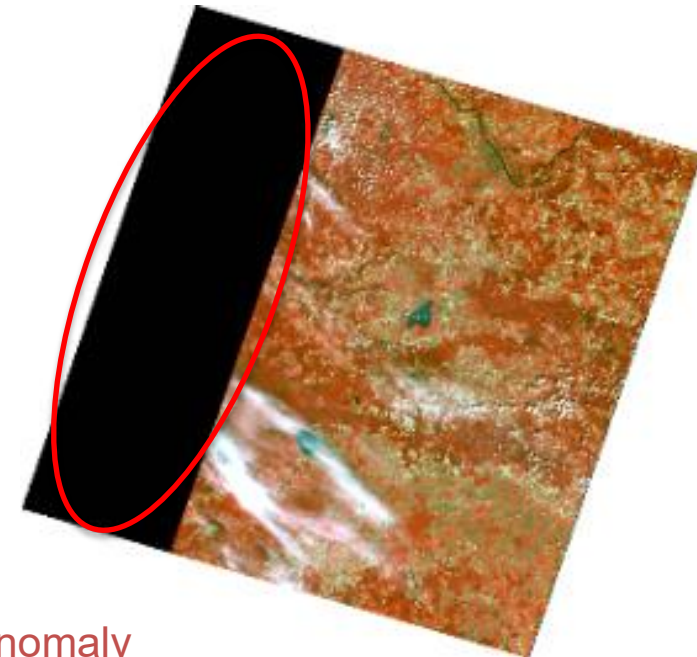
Below the table are input fields for 'Enter New Anomaly Name.' and 'Enter New Anomaly Description.', and a 'Save' button. A red arrow points from the 'Resources' tab to a metadata box below the image.

```

GROUP = L1_METADATA_FILE
GROUP = METADATA_FILE_INFO
ORIGIN = "Image courtesy of ESA"
REQUEST_ID = "0008209170000_00000 "
LANDSAT_SCENE_ID = "LM32050401982260FUI00"
ORIGINAL_FILENAME = "LM03_L1TP_205040_19820917_20190514_FUI"
FILE_DATE = 2019-05-14T10:58:42Z
STATION_ID = "FUI"
PROCESSING_SOFTWARE_VERSION = "SLAP_03.08"
DATA_CATEGORY = "NOMINAL"
END_GROUP = METADATA_FILE_INFO
GROUP = PRODUCT_METADATA
DATA_TYPE = "L1T"
ELEVATION_SOURCE = "GLS2000"
OUTPUT_FORMAT = "GEOTIFF"
EPHEMERIS_TYPE = "RESTITUTED"
SPACECRAFT_ID = "LANDSAT_3"
SENSOR_ID = "MSS"
WRS_PATH = 205
WRS_ROW = 040
DATE_ACQUIRED = 1982-09-17
    
```

AI Model & Anomaly Selection

- Key decisions to be made related to the selection of the type of AI model used by the team
 - **Convolutional Neural Network** selected
 - Similar to neuron connectivity in a human's visual cortex – popular amongst Deep Learning community for image classification
 - “**Supervised**” model type used – can be trained to detect a particular anomaly type
- Anomalies selected: Scan Start Anomaly.
 - Criteria:
 - Visible in the product image
 - Deterministic detection unfeasible

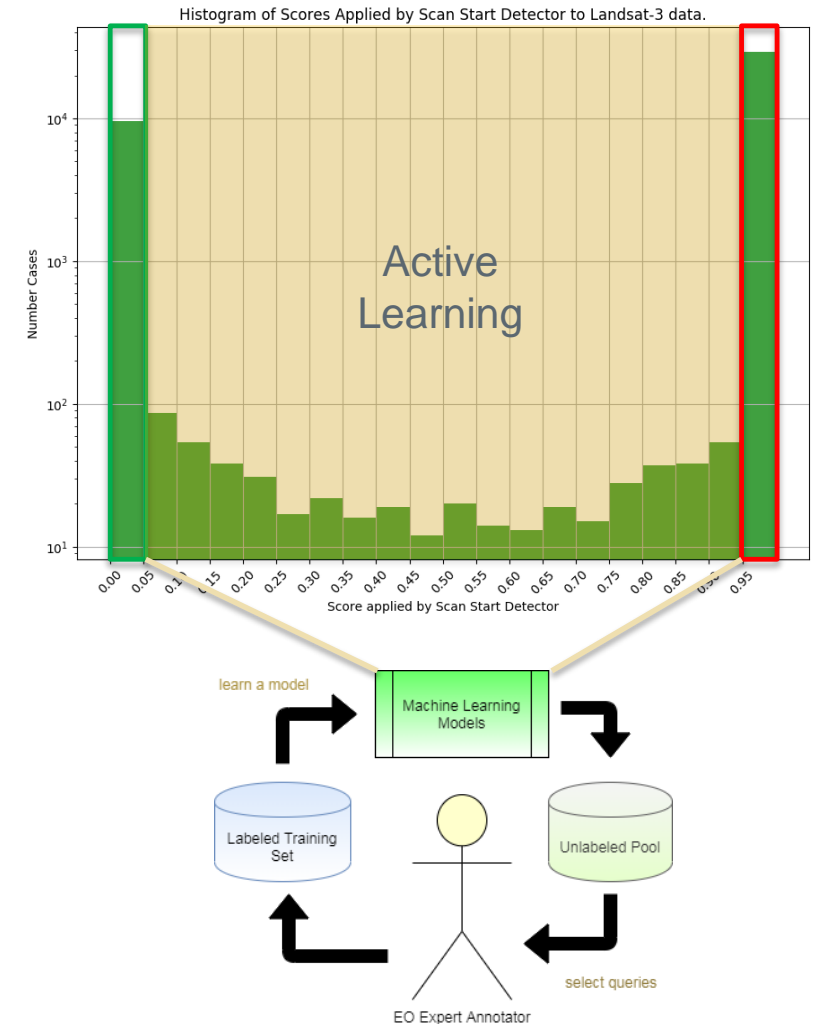


“Scan start” Anomaly



Model Refinement

- The model output is a 'soft classifier' score for each product assessed
 - Probability of a product having the anomaly rather than TRUE/FALSE
- **Active Learning** process used to improve and refine the models
- Example of results from model application (logarithmic scale) on 40'000 products
- Trained using only 25 anomalous products
- 0%-5% classified as 'No anomaly detected' / 'Negative'
- 95%-100% classified as 'Anomaly detected' / 'Positive'
- 5%-95% classified as 'Undecided'
- Undecided zone forms the basis of the Active Learning dataset
- Data is re-assessed & re-classified
- Model is re-trained and improved based on new input



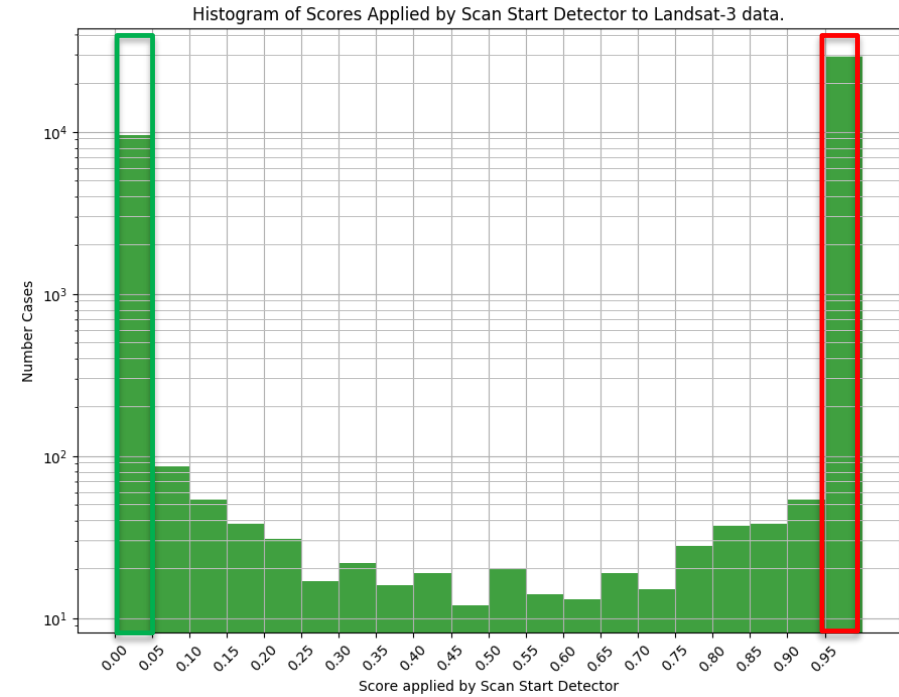


Results

- ‘Scan Start’ Anomaly
 - 39,001 Landsat-3 products analysed
 - Higher than expected Positives detected
 - Resulting from known issue prevalent towards end of mission
 - 100 samples taken across each class visually detected

	Observed	Positive	Negative
Predicted			
Positive		True positive (100)	False Positive (0)
Negative		False Negative (16)	True Negative (84)

- Model does not mix Scan Start anomaly with other missing data
- Particular cases still undetected
 - Could be improved through further training





Future Development

- Investigate potential to integrate machine learning activities into ongoing QC projects
- Expansion of techniques to include other instrument types (e.g. SAR)
- Exploration of alternative model types:
 - Unsupervised models have the potential to detect multiple anomalies with a single model, rather than having one model per anomaly
- Software tools
 - Activities to date performed on an in-house research basis
 - Software development will focus on tool integration in data processing and QC pipelines



Contacts

Andrea Melchiorre

EO Engineer

Andrea.melchiorre@telespazio.com

Kevin Halsall

Project Manager

Kevin.halsall@telespazio.com



telespazio.com



THANK YOU
FOR YOUR ATTENTION

telespazio.com

