THE UNIVERSITY OF WESTERN AUSTRALIA

Data Science Capstone Projects
CITS5553

Prof Eun-Jung Holden

July, 2022

HELP US ALL STAY HEALTHY

SIX SIMPLE TIPS

- Maintain 1.5 metres distance between yourself and others where possible
- Cough and sneeze into your elbow or a tissue (not your hands)
- Avoid shaking hands
- Put used tissues in the bin
- Wash hands with soap and warm water or use an alcohol-based hand sanitiser after you cough or sneeze
- Do not touch your face

IF YOU ARE UNWELL AND WORRIED ABOUT COVID-19:

- Call the National Coronavirus Helpline: 1800 020 080
- Call your usual GP for advice
- Call the UWA Medical Centre for advice: 6488 2118

UWA FAQs:
uwa.edu.au/coronavirus

Report COVID-19 hazards and suspected/confirmed cases via RiskWare:
uwa.edu.au/riskware

THE UNIVERSITY OF WESTERN AUSTRALIA

DCS:507757521

Unit Coordinator

- **Prof Eun-Jung Holden**

https://www.uwa.edu.au/profile/eun-jung-holden

- BSc, MSc, PhD Computer Science, UWA
- *Research areas*: Geodata science, data fusion, machine assisted geology/resource modelling, deployable data science applications
- *Research impact* : Commercialisation of three suites of geodata analytics algorithms; 3 industry driven patent applications; 2 public domain software products
- Office: **Robert Street Building Rm 1.05**
- Email: eun-jung.holden@uwa.edu.au
- Consultation time : 2-4pm **Monday**

CITS5553 Introduction

- **Outline**
  - **Unit objectives**
  - **Expected outcomes**
- **Formal class meetings**
- **Client meetings**
- **Assessments & Roadmap**
- **Seminar “Data Science Project Development & Delivery for the Minerals Industry”**
- **Projects & Groups**

Outline

- **Objective** is to provide a capstone experience where students integrate and apply knowledge and skills acquired in earlier units to a real-world data science project.
- Team based project:
  - Students work in teams of typically ~5-6 members
  - Industry/industry-relevant academic end-user client(s)
- Requirements for students:
  - Managing their project through group meetings and client meetings
  - Preparing the project proposal documentation
  - Understanding the nature of the data used in your project; choosing and implementing appropriate data analytical processes; analysing the output
  - Submitting group deliverables against milestones to clients and for unit assessment
  - Individuals are expected to perform professionally in a team-based environment and reflect on their contribution to the team.

Formal Class Meetings

- Introduction to CITS5553 (25 July, 1-2pm)
- Talk: Data Science Project Development and Delivery for the Minerals Industry (25 July, 2-3pm)

GUEST LECTURES

- 27th of July (Wednesday 11-1)
  - (F2F, recording) Deployable data science applications from a developer's perspective (Dr Daniel Wedge, UWA, 11am-12noon)
  - (F2F, recording) Collaborative development (Dr Tom Horrocks, UWA, 12noon-1pm)
- 3rd of August (Wednesday, 11-1)
  - (F2F, recording) TBC (Dr Tamara Vasey, Dr Emet Ayra, Rio Tinto, 11am-12noon)
  - (F2F, recording) Data Science in financial services (Dr Nick Lewins, Advisor, 12noon-1pm)

Dr Daniel Wedge, Senior Research Fellow, School of Earth Sciences, UWA

Dr Tom Horrocks Research Fellow, School of Earth Sciences, UWA

Dr Tamara Vasey Principal Research Lead - Digital Integrated Mine, Rio Tinto

Dr Nick Lewin Data & AI Advisor

Guest Lectures

- 10th of August (11am-1pm)
  - (F2F, recording) Ethics in AI (Dr Nin Kirkham, UWA)
- (recording) Project Team Management (Prof Sally Male, UoM)

- Some other lectures as opportunities arise

Dr Nim Kirham
Lecturer, Philosophy,
Faculty of Arts, Business,
Law and Education, UWA
“Ethical AI”

Prof Sally Male
(Recording) Director,
Teaching And Learning Lab
(Tll), The University of
Melbourne,
“Project Team Management”

Client Interactions

- Groups are to organize client interactions
  - An initial meeting to discuss the problem, expectation of the deliverables, data access, timeline to prepare the proposal document (week 1 or 2)
  - A proposal discussion meeting to propose, modify and finalise the project proposal document (before week 4 proposal submission)
  - Progress meetings (as agreed with client)
  - Final meeting (presentation) and submission of the report & codes (before week 12 final submission)

- The above are the **minimum** required interactions. Further communications and meetings may be organized if client agrees

Assessment

1. **15% Project Proposal (Group and Individual, Week 4, Friday, 19 Aug)**
2. **35% Essay on Transparent ML (Individual, Week 8, Friday, 23 Sept,)**
3. **40% Final report & codes (Groups & Individual, Week 12, Friday, 21 Oct)**
4. **10% Project presentation video (Individual, Week 12, Friday, 21 Oct)**

- Late submission penalty: 5% per 24 hours (as per UWA rules)

Timeline

Week 1
- **Information session**
- **Guest Lecture**: Deployable Data Science Applications from a Developer's Perspective
- **Guest Lecture**: Collaborative Development of Data Science Application using GitHub

Week 2 &3
- **Guest Lecture**: Day of a Data Scientist
- **Guest Lecture:** Ethical AI

Week 1-4
1. Group project proposal (group agreed by client; and individual plan)

Week 5-8
2. Essay on Transparent ML (individual)

Week 1-12
3. Final group report & codes (group, individual contribution)

Week 1-12
4. Final presentation video (individual)

1. PROJECT PROPOSAL (15%)

- **Group Project Proposal (Group, Max 3 pages without Appendix)**
    - PROJECT TITLE
    - AIM
    - BACKGROUND - Understanding the data & the current data interpretation practice
    - VALUE PROPOSITION – why use data science
        - Understanding bottleneck/challenges in client's data interpretation
        - Have a clear value proposition (efficiency, consistency/reproduciblity, cost-saving etc)
    - DELIVERABLES
        - Expected outcomes of the project
            - Data analytics & visualisation outputs in presentation, reports and codes
    - METHODS
        - Preliminary interrogation of the data
        - Choice of different data exploration and analytical techniques
        - How to make the process understandable and the output usable (e.g. parameter control, incorporating domain knowledge, communicating output uncertainty)
        - Visualisation to communicate the processes/results
    - PROJECT DELIEVERY TIMELINE (Team time commitment)
    - PROJECT MANAGEMENT TIMELINE (next slide)
- **APPENDIX**
    - Individual plan and reflection on their contribution to the project for each member (**max 400 words per member**)

Project Management

- Choose the team leader (a point of contact for your client)
- Discuss the aims; the scope of the project; and data (understanding the characteristics)
- Literature reviews on potential data science methods and identify different approaches (e.g. statistical vs ML based clustering, different classifiers, etc.) suitable for the project
- Choose the programming language, libraries, and the platform that are to be used
- Team members decide on implementing & testing different approaches & their output visualisation
    - Individuals are responsible for their testing of the chosen approach
    - The group will work together to compare the outputs from different approaches for clients
- The report (executive summary) from the team needs to incorporate members' work and a combined assessment of the applied data science methods
- NEED to build a PROJECT MANAGEMENT TIMELINE including the tasks for individuals and their responsibilities, including the role of the team leader

- "A Typical Sprint: Play-by-Play" by Ian Mitchell,
- https://www.scrum.org/resources/blog/typical-sprint-play-play

2. ESSAY (Individual, 35%)

Max 3000 words

- The topic of transparency and explainability of data science solutions
  - Literature reviews
  - How these considerations are to be applied to your project

Example from 2021 Refs

- Ribeiro, M.T., Singh, S. and Guestrin, C., 2016, “Why Should I Trust You?” Explaining the Predictions of Any Classifier, Proceedings of KDD, San Francisco, USA
- Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* **1,** 206–215 (2019). https://doi.org/10.1038/s42256-019-0048-x

…This year’s refs will be released later.

3. FINAL REPORT (Group & Individual, 40%)

- **GROUP REPORT For Client (Max 7 pages including figures)**
  - Introduction
    - Problem; current practice; the specific contribution of the project; and outcomes
  - Methods & Results
    - Summary of different analytical techniques and results
    - Considerations on interpretability and/or transparency
  - Conclusion
    - Comparison of the outputs from applying different techniques
    - Summary
  - GitHub repository link for the group work

- **APPENDIX**
  - Individual report on their contribution (**3-5 pages including figures per person)**
  - Individual GitHub repository (within group repository)
  - Individual reflection on team work experience & Challenges experienced

4. PRESENTATION VIDEO (Individual, 10%)

5 mins

- Introduction to the problem and data
- Your contribution to the project
- Transparency/actionability of the methods
    - What were the considerations (from your essay)
    - How your solution can assist decision for end users

Data Science Project Development & Delivery for the Minerals Industry

Prof. Eun-Jung Holden

Director, UWA Data Institute

Leader, Centre for Data-driven Geoscience

THE UNIVERSITY OF WESTERN AUSTRALIA

UWA DATA INSTITUTE

Centre for Data-driven Geoscience

Geology & Data

Geology is important for

- *Finding locations of mineral potential*
- *Modelling resource*
- *Safe and effective mining and extraction*

Minerals industry uses:

- *Field observations*

  Mapping of structures, lithology, stratigraphy, reports etc

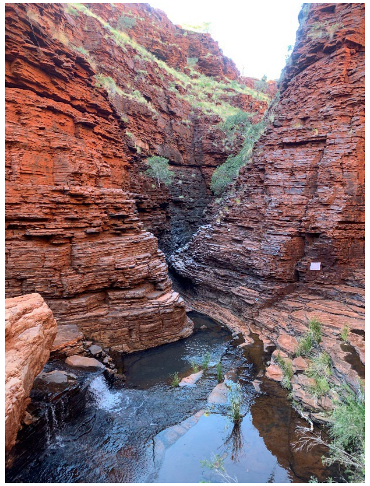
- *Data from rock samples (surface or drillhole)*

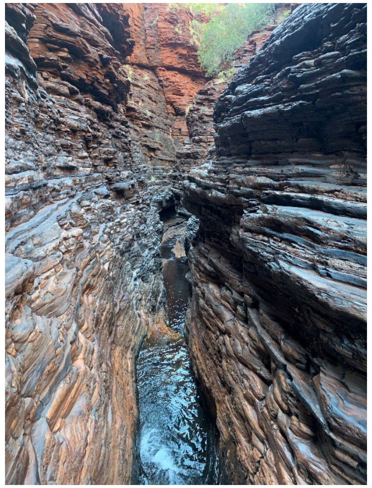
  Mineralogy, lithology, petrology data through sample/rock chips/core interpretation, thin section analysis, multi-element assays, and/or spectral analysis, petrophysical properties etc

- *Data from remotely sensed techniques*

  Ground surveys (gravity, magnetic susceptibility, seismic, etc); drillhole data collection (televiewer images, wireline measurements); airborne surveys; satellites

- *Geological Documents*





Data, Decisions, Machine Augmentation

DATA
(Structured, unstructured)

Machine Augmentation

INFORMATION

- refined form of data

KNOWLEDGE

- collective information (e.g. concepts) that are acquired in the field of expertise

WISDOM

DECISION

Industry Challenges:

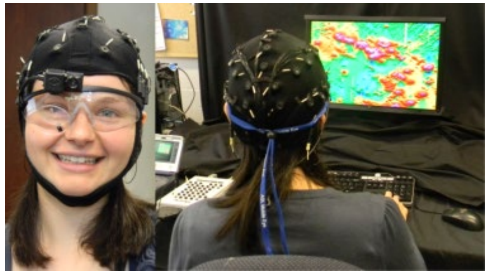
- *Ensuring return on investment on costly data collection*
- *Large volume of data & lack of resources creating industry bottleneck*
- *Diverse skillsets and software required to process and analyse different types of data*
- *Human bias & inconsistency (within and across individuals) leading to non-reproducible outcomes*

Machine Augmentation can assist using:

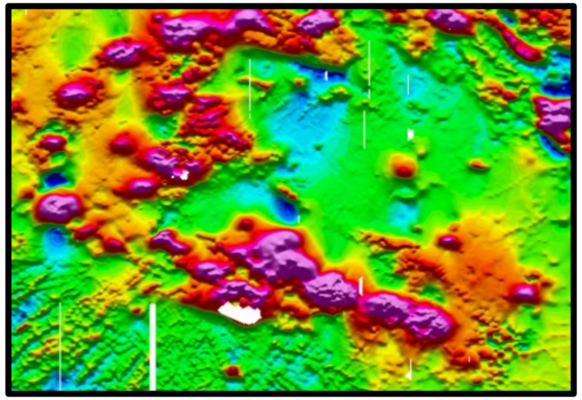
- *Pattern enhancement or recognition techniques (signal/image analysis techniques)*
- *Prediction, regression or clustering techniques (machine learning or statistical analysis)*
- *AI techniques (natural language processing, computer vision)*
- *Visual analytics & human computer interaction*

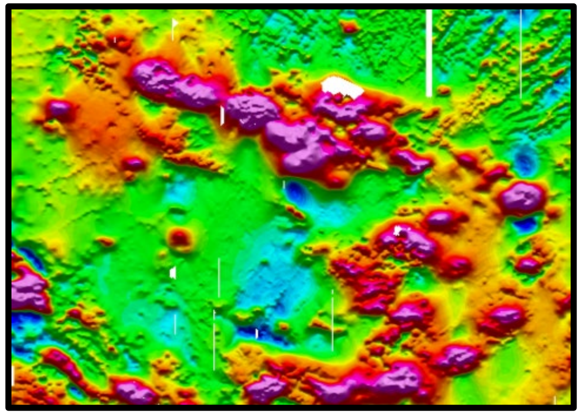
Yathunanthan Sivarajah

Pattern Recognition



- Porphyry target detection is ***highly variable*** amongst and within individuals
- ***Visual attention*** is drawn to areas of ***high contrast*** in colour, intensity, orientation





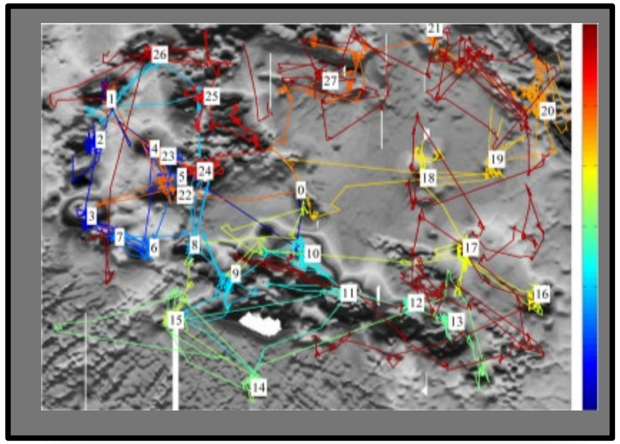
Data Property of Barrick Gold

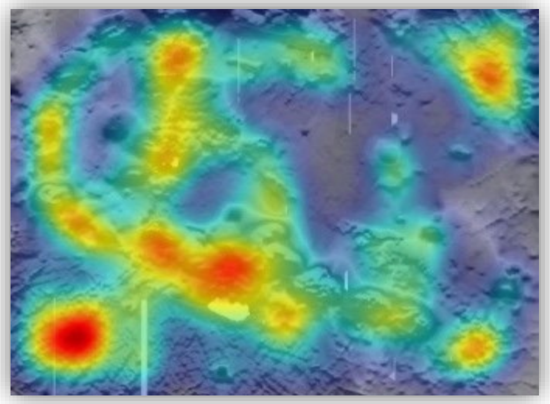
21 Ground Truth Targets

F-measure
= 2 (R + P) / (R x P)

Recall (R) is the proportion of the detected ground truth targets in all ground truth targets(tp/(tp+fn))

Precision (P) is the proportion of the detected ground truth targets in all detected targets (tp/(tp+fp)).





Y Sivarajah, *EJ Holden, R Togneri, M Dentith (2013) Identifying effective interpretation methods for magnetic data by profiling and analyzing human data interactions, Interpretation 1 (1), T45-T55*

Y Sivarajah, *EJ Holden, R Togneri, G Price, T Tan (2014) Quantifying target spotting performances with complex geoscientific imagery using ERP P300 responses, International Journal of Human-Computer Studies 72 (3), 275-283*

Y Sivarajah, *EJ Holden, R Togneri, M Dentith, M Lindsay (2015) Visual saliency and potential field data enhancements: Where is your attention drawn? Interpretation 2 (4), SJ9-SJ21*

X-ray computed tomography (CT) scan of a human lung & the task was to find bright white cancer nodules (Drew, 2013)

Drew, T., Võ, M.L.H, and Wolfe, J.M., **The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers,** Psychological Science, September 2013 24: 1848-1853

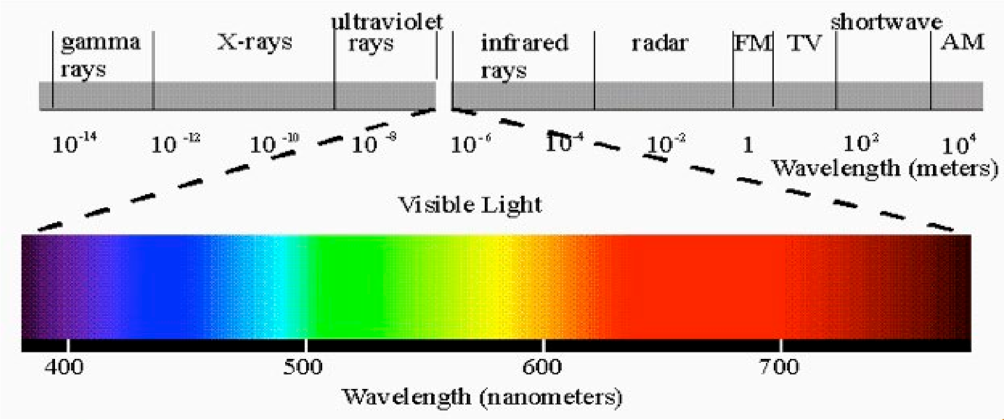
- Dancing gorilla was hidden in 239/1000 scans – 4/24 radiologists spotted the gorilla

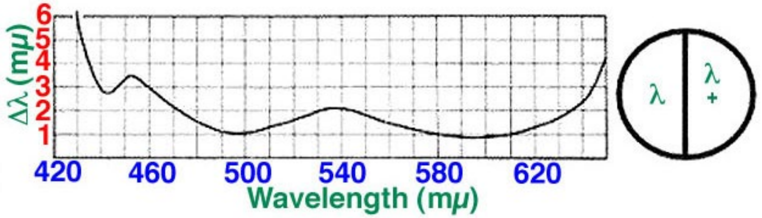
Inattentional Blindness – not likely to find features which are not expected

Human Colour Perception

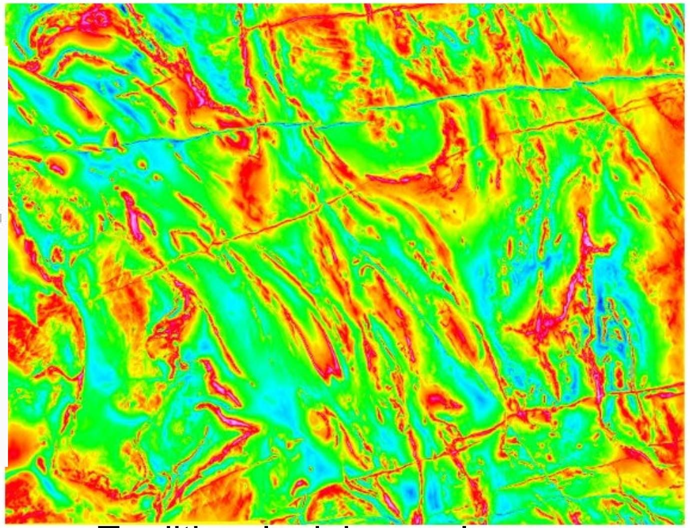
Dr Peter Kovesi:
https://colorcet.com

- Within visible light portion of electromagnetic radiation, our perceptual contrast varies across the spectrum wavelength range
  - Potentially see false features at points of low perceptual contrast
- Perceptually uniform colourmap is available

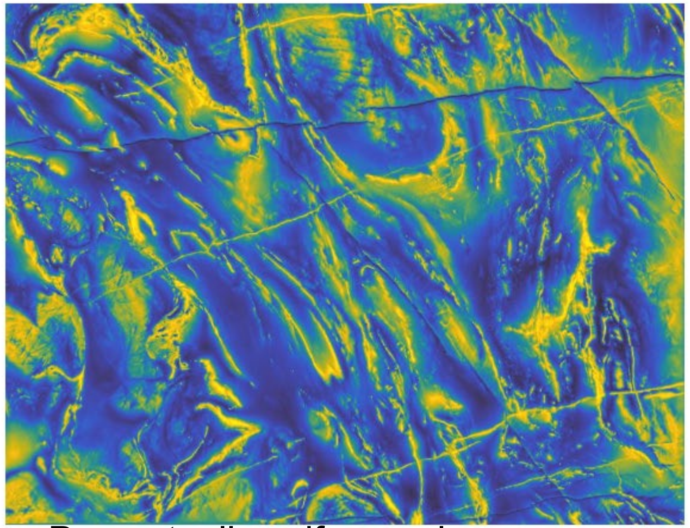




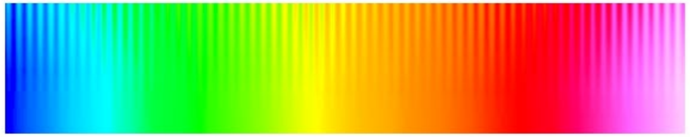
Mean wavelength discrimination curve (From Davson, H., The Eye, Vol 2. London Academic Press, 1962),
Source:
http://webvision.med.utah.edu/imageswv/KallColor13.jpg



Traditional rainbow colourmap

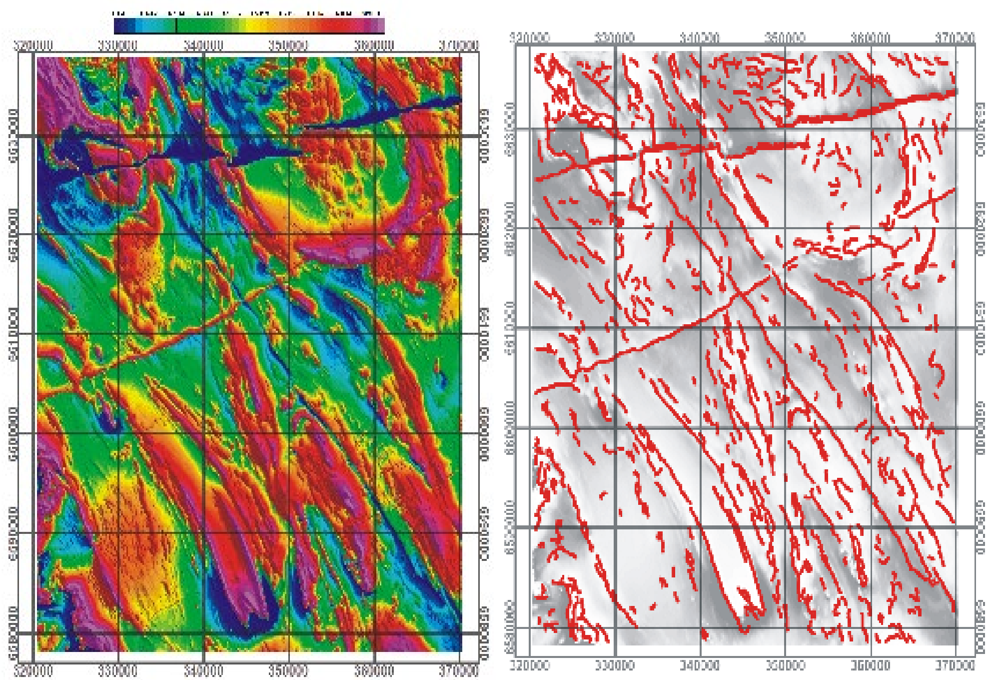


Perceptually uniform colourmap



Test image consisting of a sine wave superimposed on a ramp function. The amplitude of the sine wave is progressively reduced to zero at the bottom of the image

Machine-driven Interpretation



(Commercialised in 2010) CET Grid Analysis extension for Seequent (previously Geosoft) Oasis Montaj

Structural interpretation of magnetic data

- Structures appear as discontinuities within data
- Automated analysis is fast and repeatable
- False positives from automated analysis require human effort to remove/correct
- Structure mapping is more than finding lineaments
  - Different types of structures/features
  - Chronological order
  - 3D orientation

For complex interpretation tasks, automated analysis can assist but not always possible to replace human interpretation

In 2016, we released Integrated Exploration Platform – Geological Survey of WA product (https://www.waexplorationplatform.wa.edu.au/) including a workflow where automated analysis is used to interactively guide interpretation & provide quantitative confidence of manually interpreted lines.

Human, Machine, Data & Rocks

– **HUMAN:** Highly variable interpretation outcomes but hold domain knowledge

– **MACHINE**: Fast, objective and reproducible outcomes but building a data science application to replicate human decision is difficult for domain specific and complex tasks

– **DATA**: Large volumes of different modality data at different resolution; Data rich but information poor

– **ROCKS**: Geology is complex & difficult to model

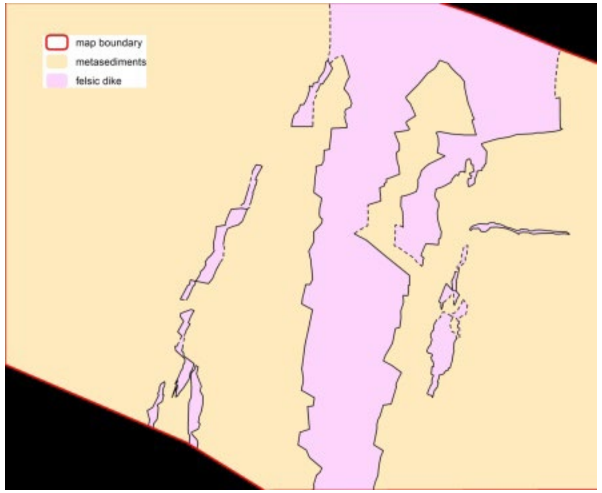
My research (Centre for Data-driven Geoscience) aims to build interpretable and actionable data science applications to assist the geological interpretation of diverse types of geodata – Support MACHINE AUGMENTED INTERPRETATION

Semi-automated Lithology Mapping

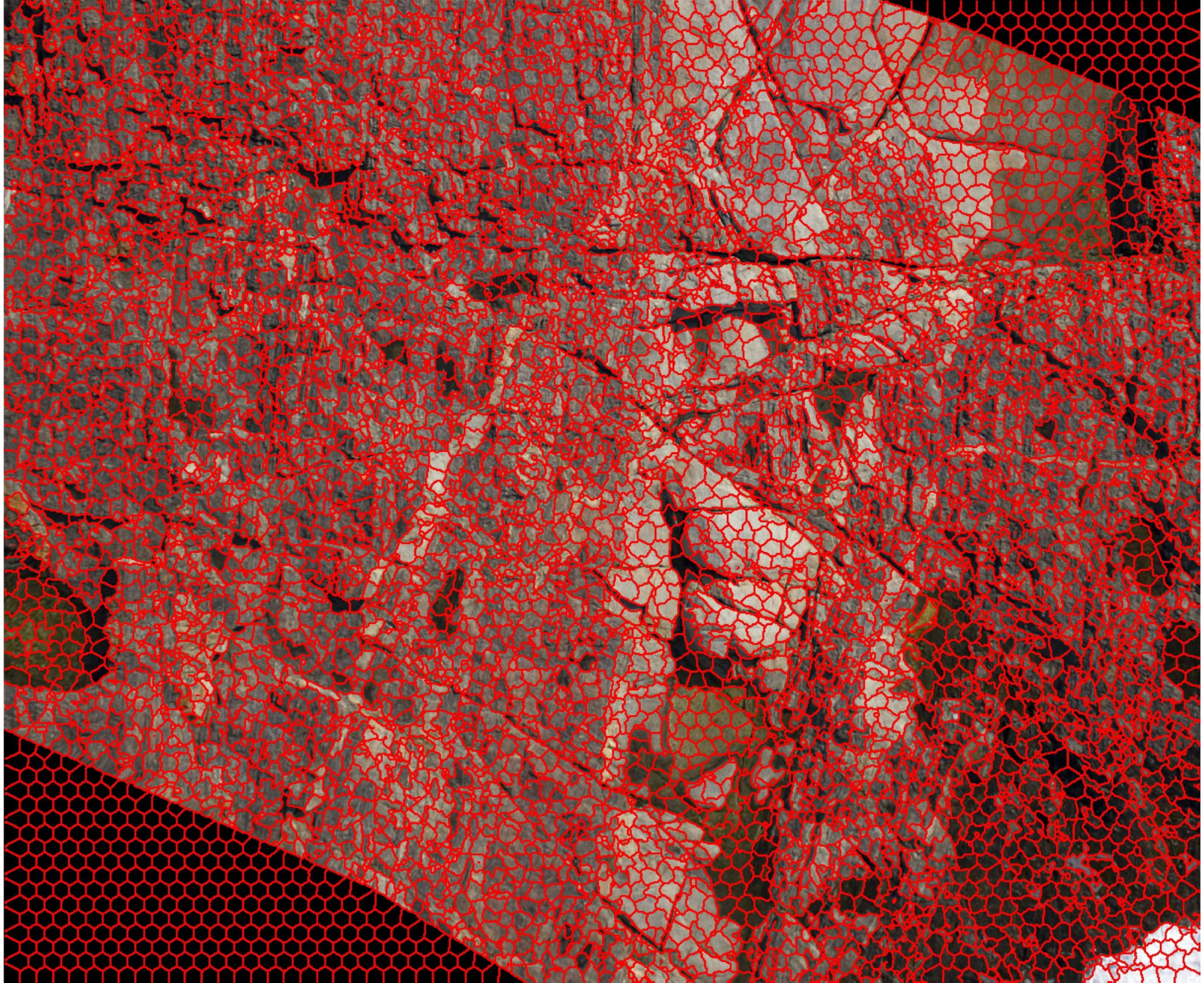
Student:
Vasuki
Yathunanthan

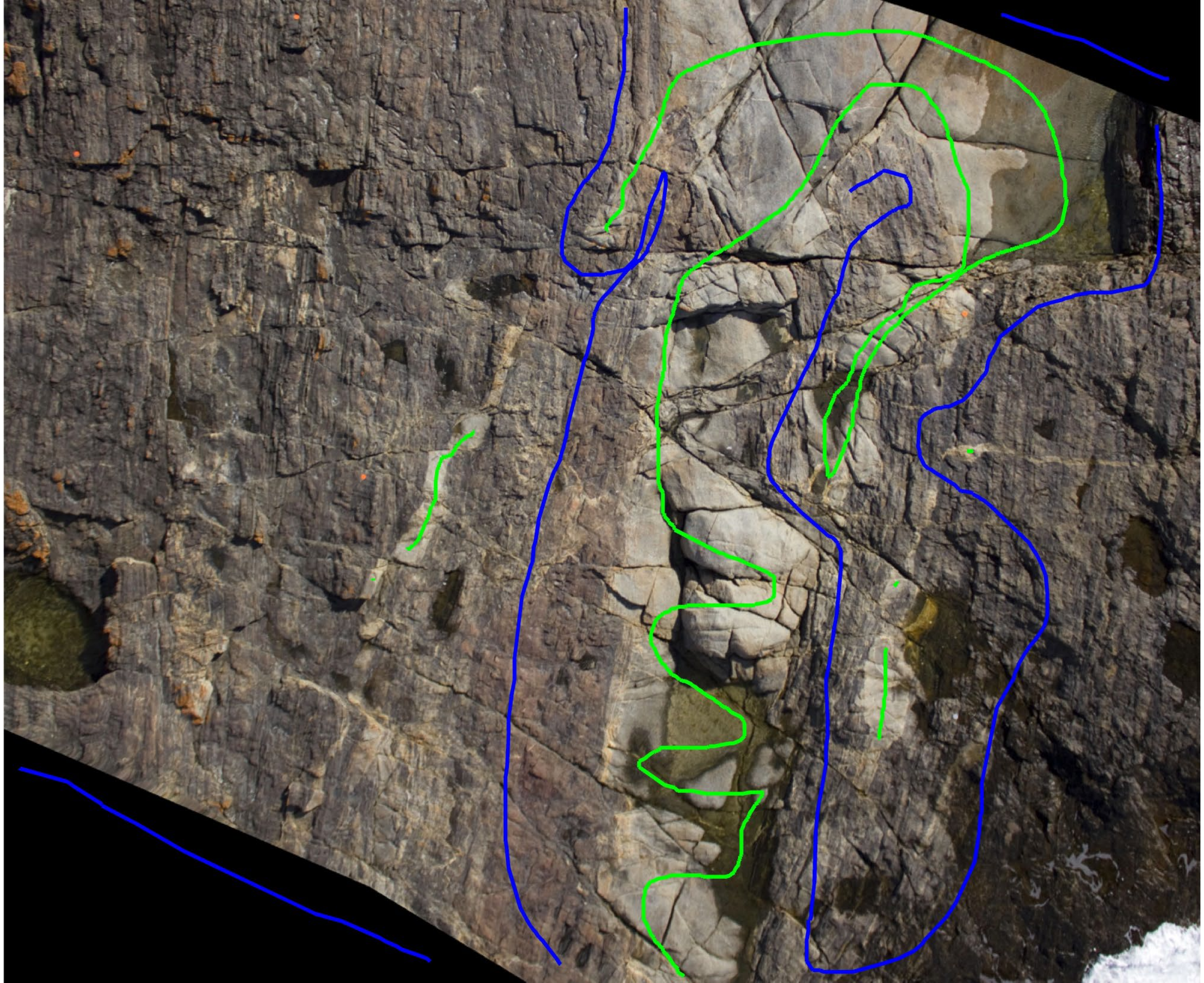
Oktokopter fitted with Canon 550D

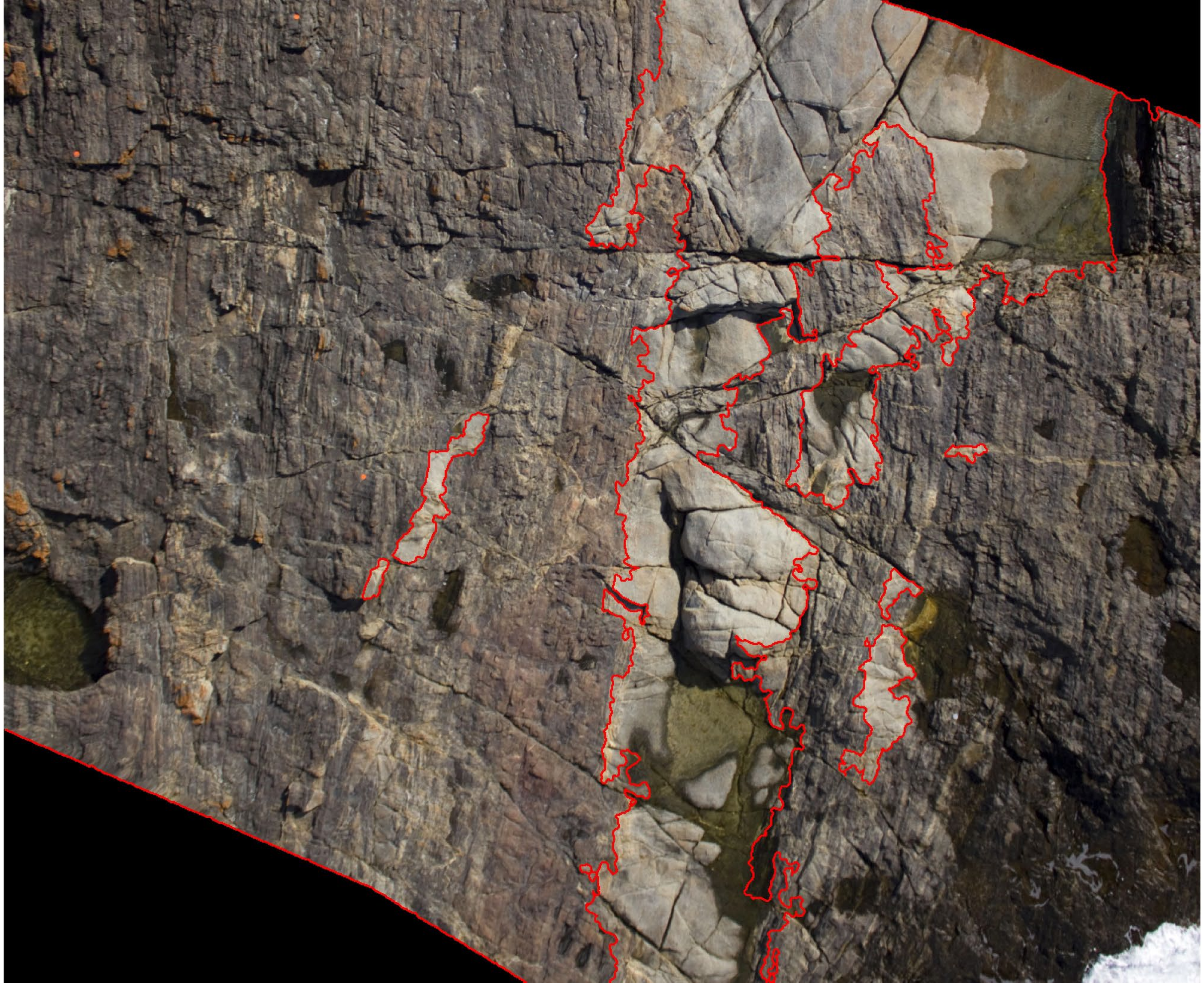
Survey was flown by Darren Turner & Arko Lucieer from University of Tasmania

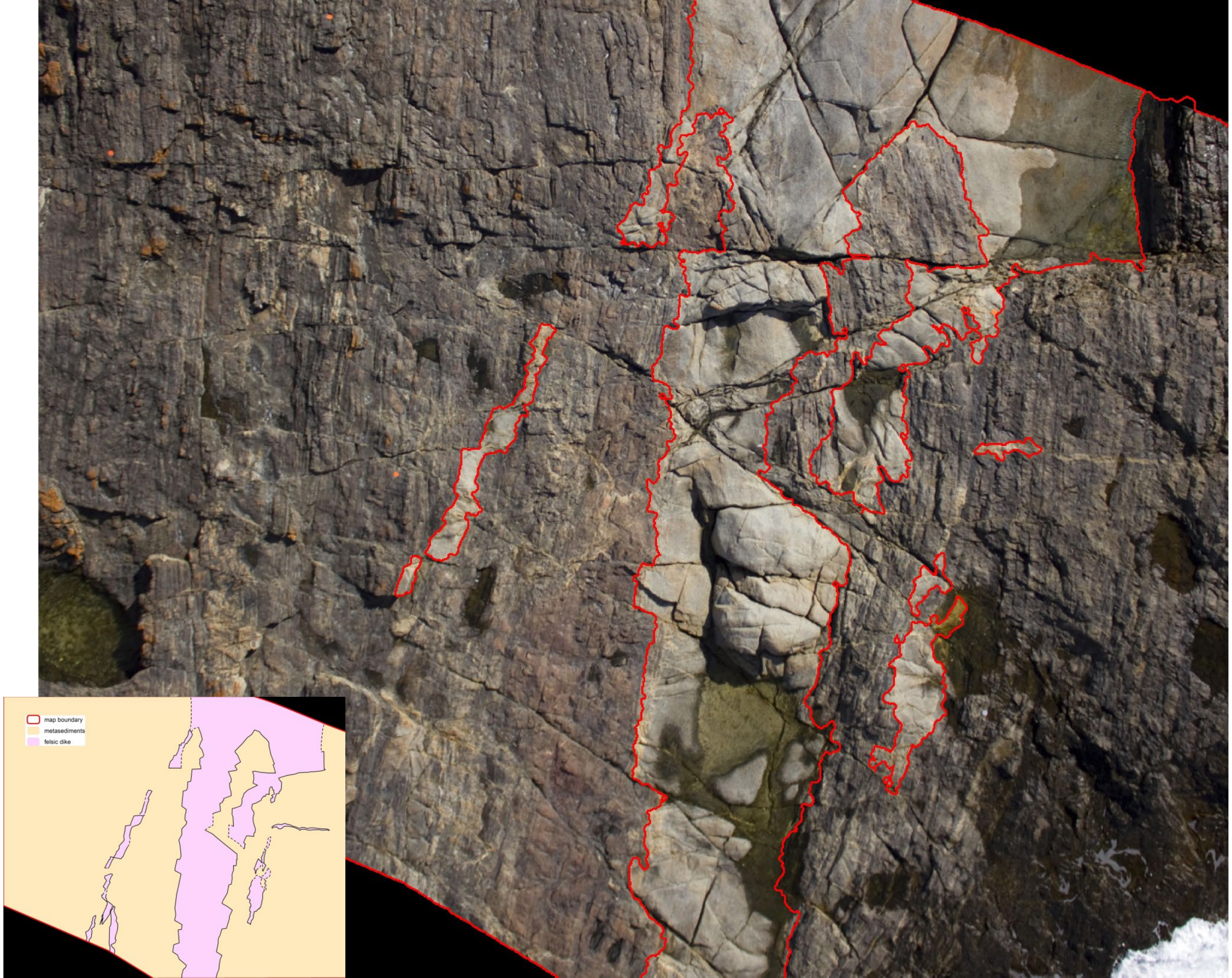


Manual interpretation of Steve Micklethwaite

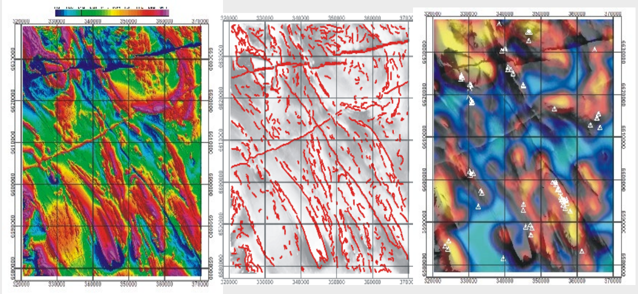




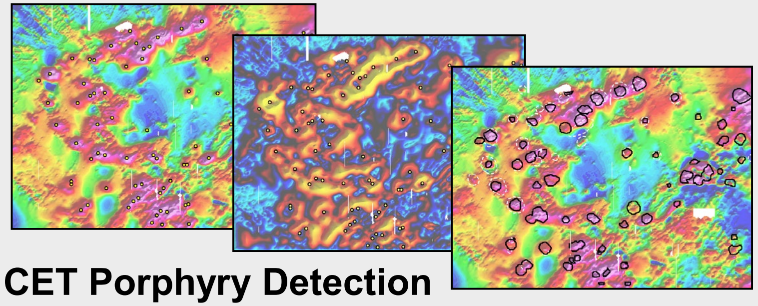



map boundary
metasediments
felsic dike

Deploying Data Science Applications

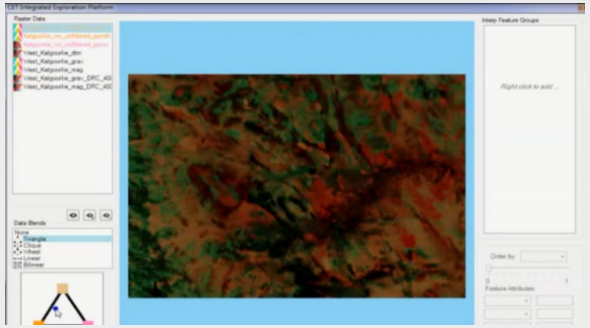


CET Grid Analysis



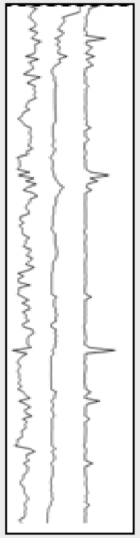
CET Porphyry Detection
(Barrick Gold)

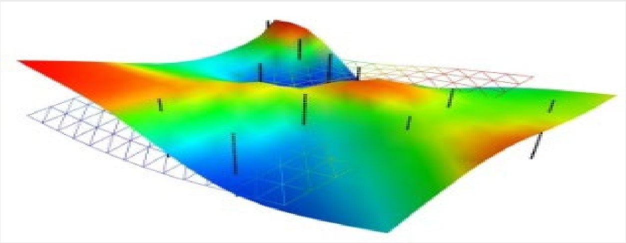
Commercialised as
Geosoft/Seequent Oasis Montaj extensions (2010-ongoing)

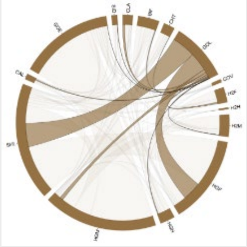


Integrated Exploration Platform (IEP)
(GSWA-ARC linkage)

Public domain GSWA product (2017)

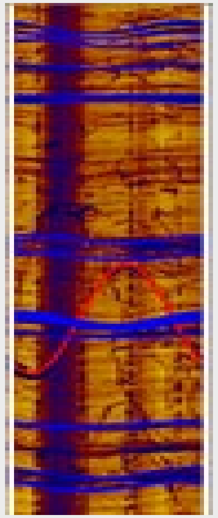






Machine learning tools to improve the analysis and integration of drillhole and other data to improve ore body knowledge *(Rio Tinto Pty Ltd)*

Three PATENT Applications :
A Method And System For Validating Logging Data For A Mineral Sample (AVA, 2018); A Method And System For Sample Classification (MADI, 2019); Provisional patent application on A Method And System For Logging Data For A Mineral Sample (Objective Logging, 2021)



Televiewer image analysis
(Rio Tinto Pty Ltd)

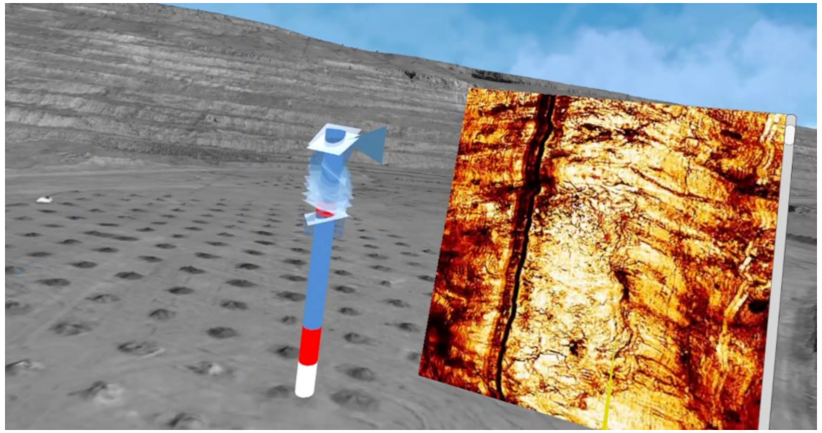
Commercialised into ALT's WellCAD Image & Structure module (2015-ongoing)

Deployable Data Science Applications

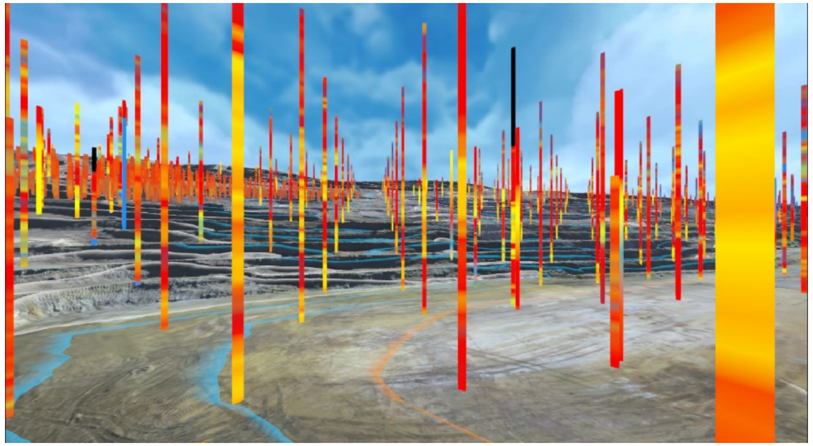
1. Identifying ‘high-value’ problem
2. Understanding data, domain, how they are interpreted
3. Responsible use of data science to build ‘trust’ in machine augmentation
   - QA/QC of false positives/negative from automated solution take long time to clean up
   - Algorithmic Transparency (e.g. how it works, impact of parameters, input features vs prediction)
   - Communicating uncertainty of automated solution (e.g. input & output variability)
   - Extensive evaluation
4. Addressing ‘Usability’ issue
   - Consider how machine augmentation gets integrated into practitioners’ workflow & current software
   - Transferability of the solution (e.g. ML model) to different environment
   - Dealing with data access/warehousing, when things run (triggers), model updates & legacy results etc.

Case Studies

***Case study 1**: Structural Analysis using televiewer images*



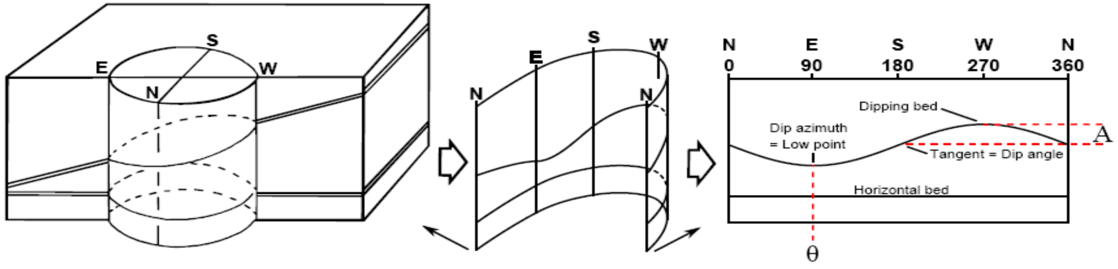
***Case study 2**: Machine Assisted Stratigraphy Interpretation using multi-modal drill hole data*



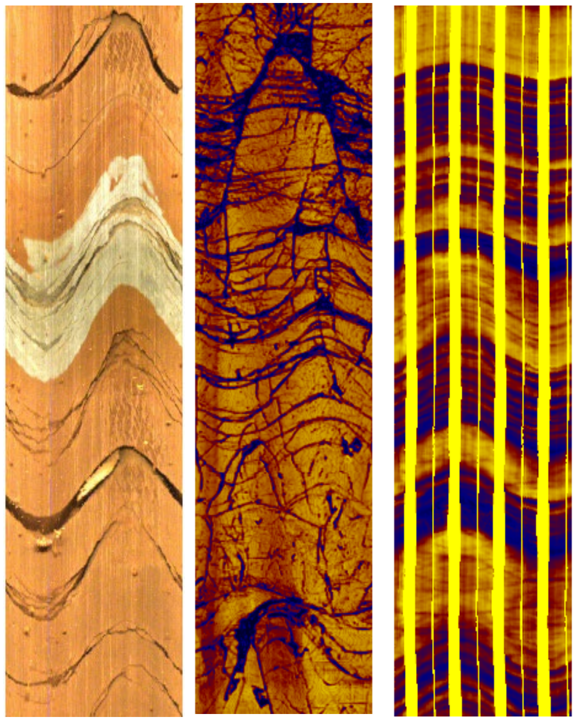
Sponsored by
Rio Tinto Pty Ltd

Dr Daniel Wedge

Case Study 1:
Televiewer Structure Analysis



Malcolm Rider, The Geological Interpretation of Well Logs.



OTV
(Optical)

ATV
(Acoustic)

FMI
(Electrical resistivity)

- Borehole structure analysis is important for petroleum (understanding fluid flow, FMI); mineral exploration (structurally controlled mineralisation, OTV, ATV); mine site design (structural integrity, OTV, ATV)
- Sponsoring company had hundreds of kilometres of ATV &/ OTV televiewer data to analyse a year creating a bottleneck in industry workflow.
- Industry sponsorship (2010-2013) to develop automated structure analysis methods

Televiewer Image Analysis

- *False positive/negatives*
- *Difference in user groups needs*
- *Interpretation needs to identify structure types*

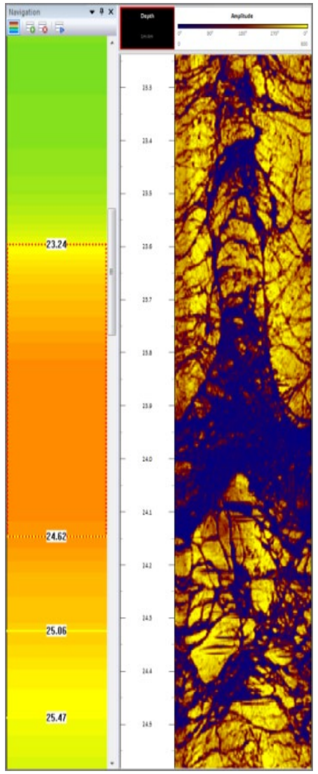
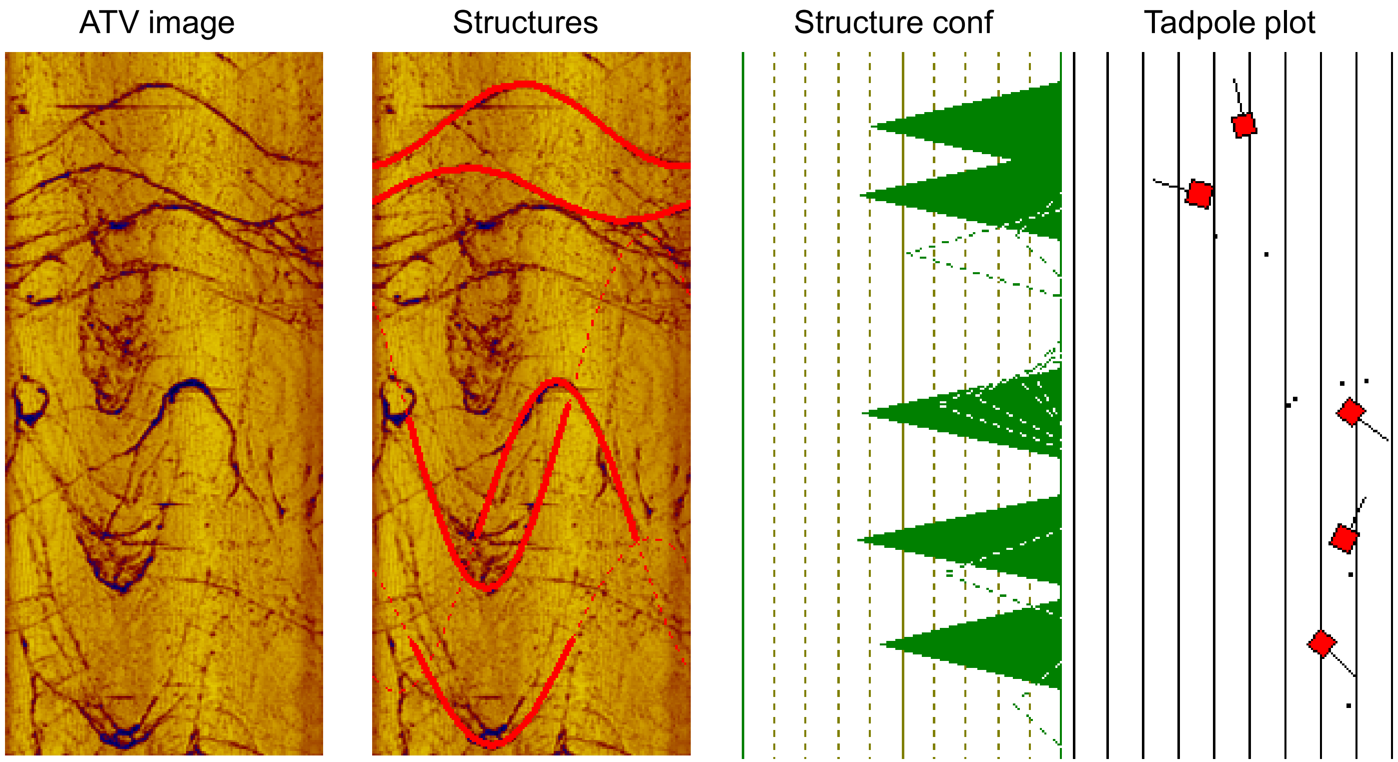


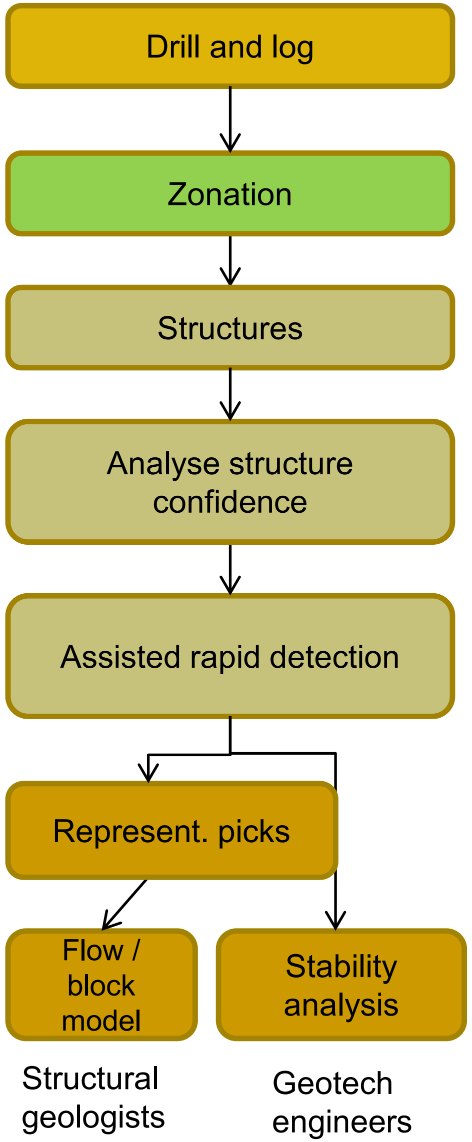
Image complexity zoning algorithms (~RQD, geological domaining)



Sinusoid detection and structure orientation estimation algorithms

Wedge, D., Holden, E. J., Dentith, M. & Spadaccini, "Fast and objective detection and analysis of structures in downhole images", 2017, Journal of Applied Geophysics. 144, p. 157-172

Structure interpretation workflow



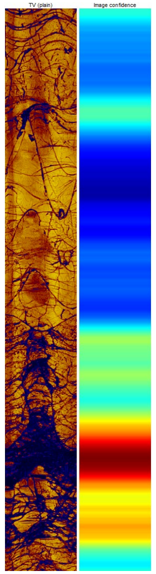
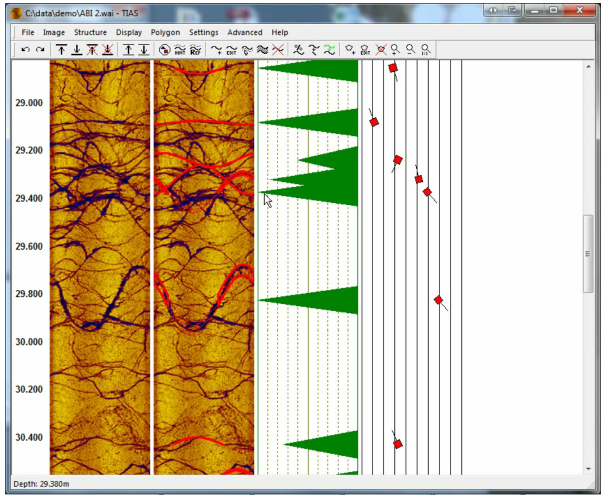
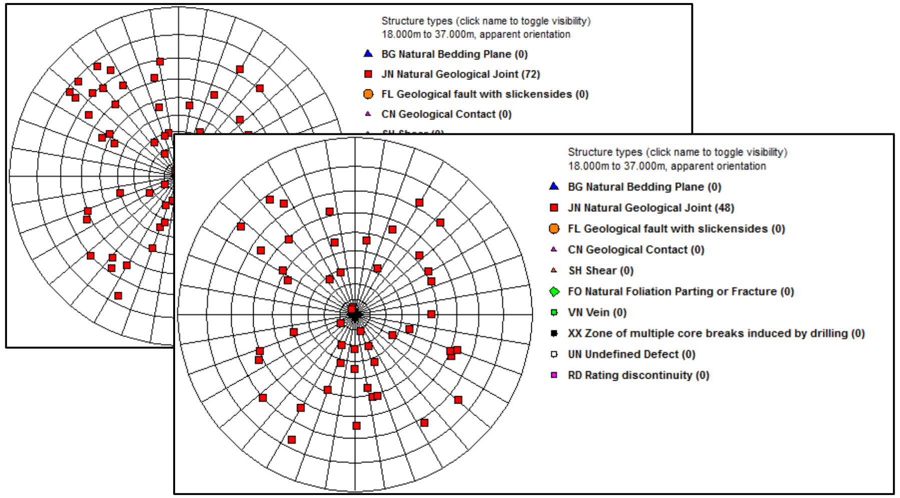


Image texture based zonation (related to separating zones of different structural complexity, imaging conditions impacting **confidence *on automated picking***)



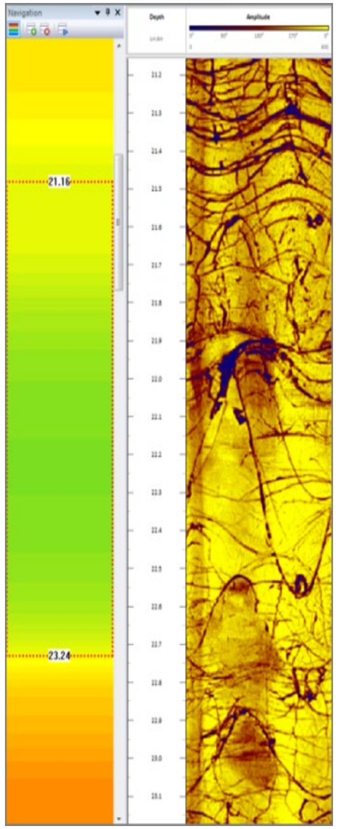
- Automatically picking sinusoids
- ***Fast QA/QC for false positives/negatives***
  - *Interactive* selection of auto picked structures based on their confidence
  - *Interactive visualisation* to allow rapid detection of similar structures, manual modification etc

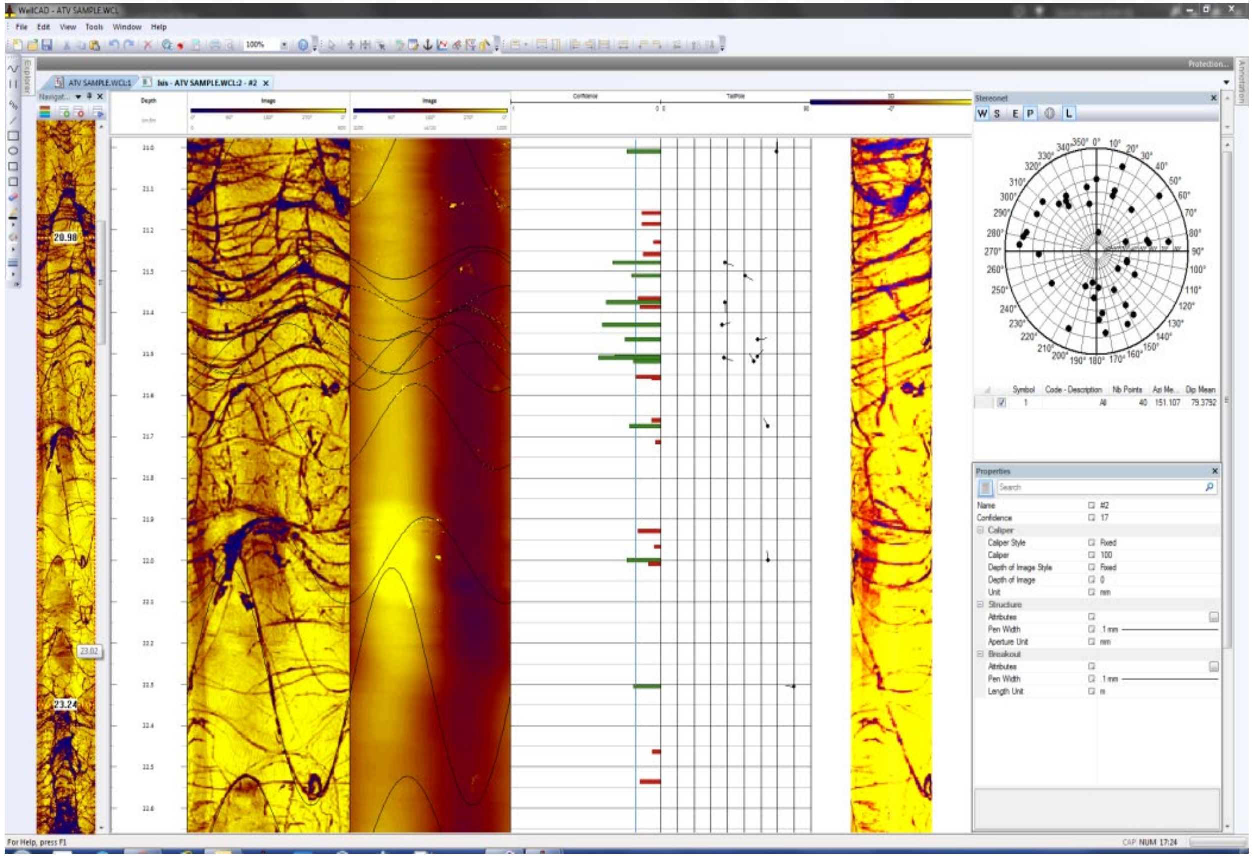


- ***Meet the needs of two different user groups***:
  - Interactive selection of representative picks based on orientation, depth separation and structure confidence
- ***Augmenting structure classification***
  - Interactive (re)classification of different structure types

Televiewer Structure Analysis

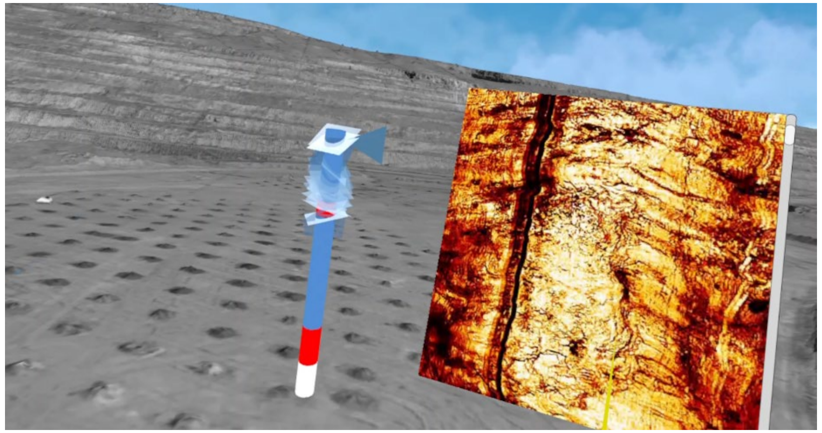
- Research & development in collaboration with Rio Tinto (2010-2013)
- **Commercialisation**: Integrated into Image & Structure Interpretation Module for WellCAD (v5.1) by Advanced Logic Technology (ALT), 2015
- Significant uptake by the resource industry globally



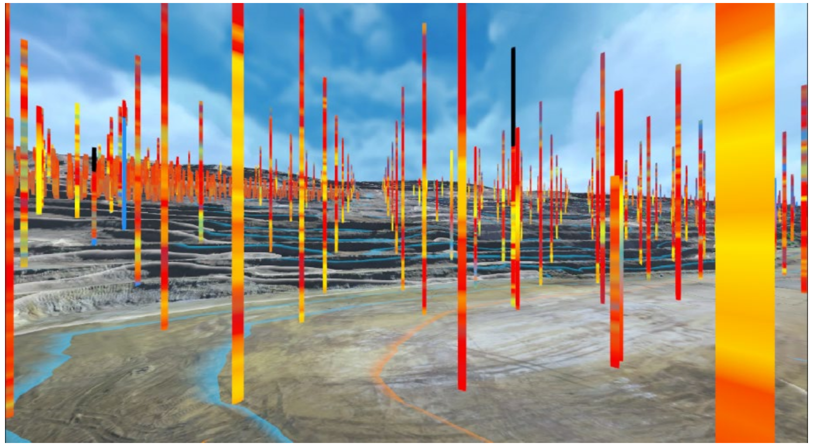


Industry Collaboration Case Studies

***Case study 1**: Structural Analysis using televiewer images*



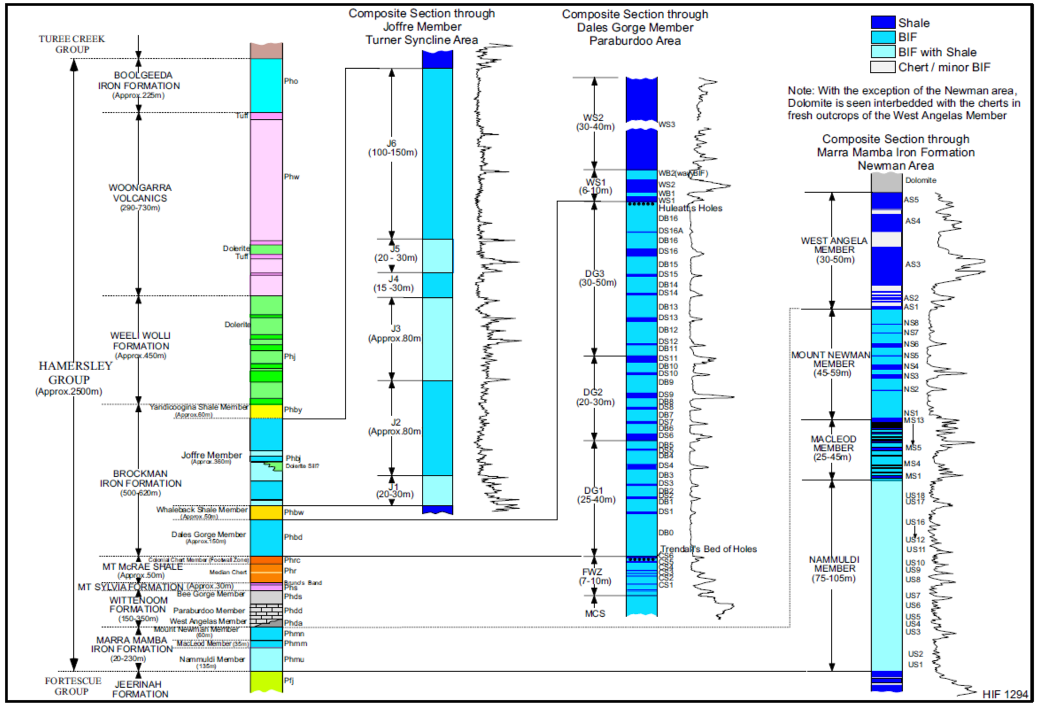
***Case study 2**: Machine Assisted Stratigraphy Interpretation using multi-modal drill hole data*



Sponsored by
Rio Tinto Pty Ltd

Dr Daniel Wedge

MADI for Resource Evaluation



Stratigraphic Columns for the Hamersley Group

- **Goal**: Identify strand units (stratigraphic units or subunits) within the drillhole in Pilbara for iron ore operation
- **Why**: Identifying where you are in the stratigraphy for the Hamersley group is important for block modelling (containing all the grade and material information)
- **Data**:
  - Multi-element assays (2m intervals)
  - Geological logging (2m intervals)
    - from RC chips logged at the rig and later **validated** against assays
  - Downhole geophysics (10cm intervals)
    - gamma, mag sus, density

***UWA-RTIO PATENT APPLICATION**, 2019, "A Method And System For Sample Classification" (PCT/AU2019/051206) by **Wedge, D**, Hartley, O., McMickan, A., Green, T. and Holden, E-J,*

***PUBLICATION**: "Machine Assisted Drillhole Interpretation of Iron Ore Resource Evaluation Holes in the Pilbara", 2019, **Wedge, D**, Hartley, O., McMickan, A., Green, T. and Holden, E-J, Ore Geology Reviews*

Improve Quality of Input by AVA (Auto Valiation Assistant)
UWA-RTIO PATENT APPLICATION: “A Method And System For Validating Logging Data For A Mineral Sample” (WO 2018/136998 A1) by Wedge, D., Paine, M., Lewan, A., Holden, E-J. & Green, T., 2018, Patent Cooperation Treaty
PUBLICATION: Wedge, D., Lewan, A., Paine, M., Holden, E. J. & Green, T., 2018, “A data mining approach to validating drill hole logging data in Pilbara iron ore exploration”, Economic Geology. 113, 4, pp. 961-972

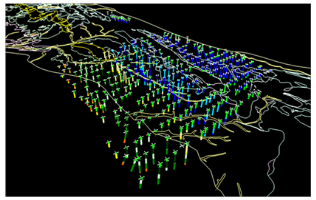
Machine Assisted Drillhole Interpretation (MADI)

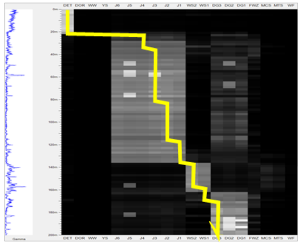
Gamma | Geological logging | Assays

CNN Classifiers

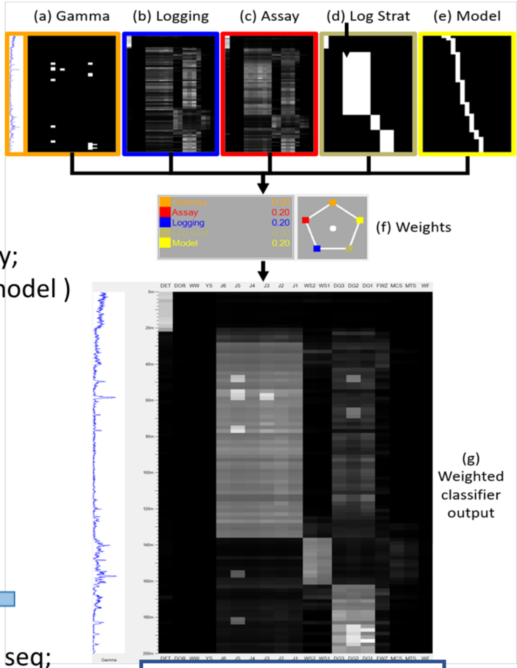
Classifiers to analyse multi-modality data @ different resolutions

Final Interpretation





Interactive Visualisation



(+ Logged Stratigraphy; Historical model)

(+ Strand seq; Strand thickness; Structures)

Optimisation

Further considerations on ***building ‘trust’ & making machine augmentation ‘useable’***

- MADI ML models were trained with ~50K intervals; tested with 11428 intervals with > 85% match with expert results
- ***Transferability*** of ML models
- ***Communication*** through interactive visualisation

Augmenting final decision through spatial visualisation

SEAMLESS Integration of Machine Augmentation in interpretation workflow

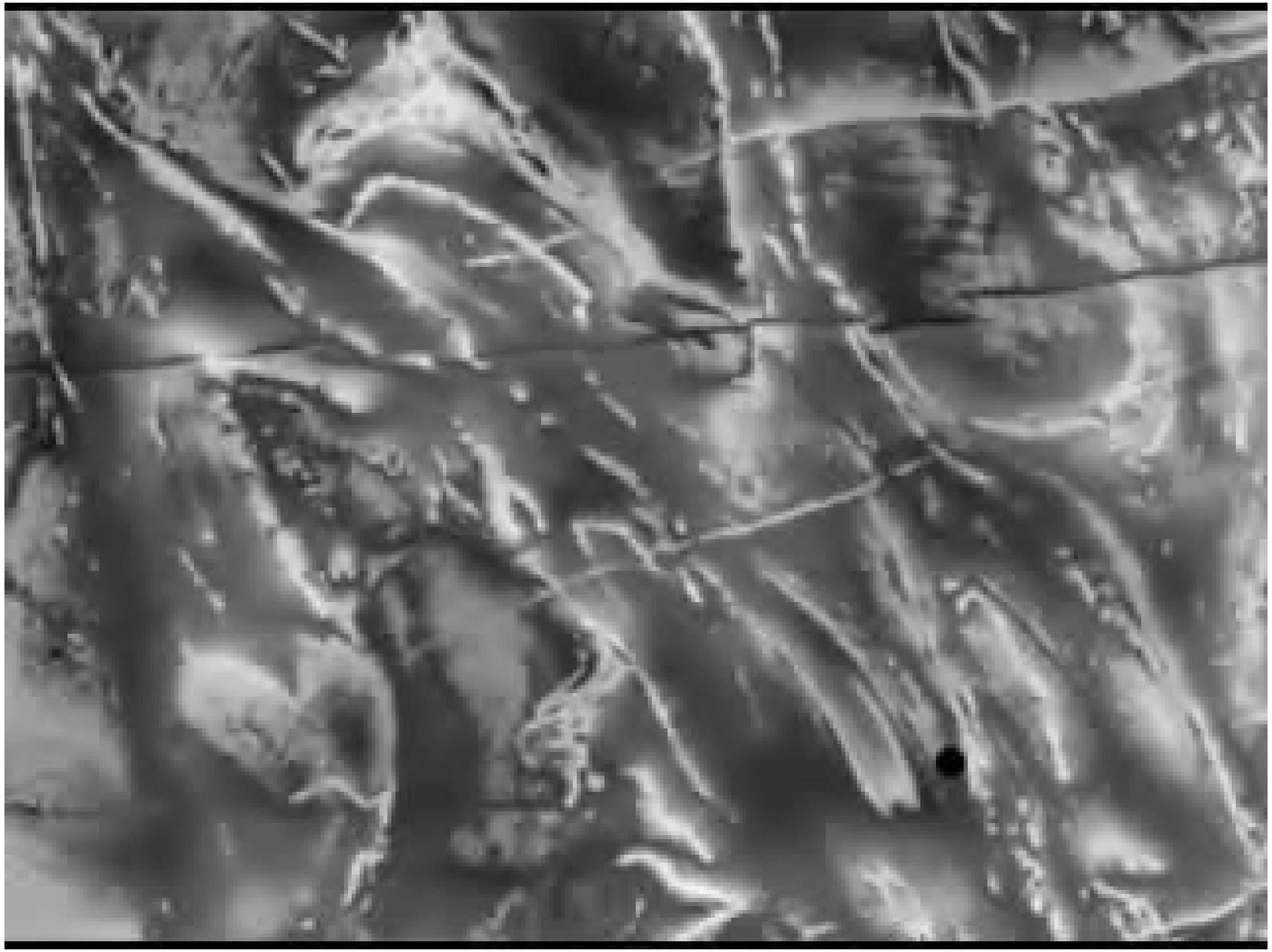
- Interactive visualisation for user control
- Integrate auxiliary data & domain knowledge to achieve geologically feasible solutions

Visual Analytics & Human Data Interaction

Dr. Peter Kovesi

Gravity *Magnetics*

Frequency cutoffs



- Interactive visualisation is important to engage user in exploring data; or in communicating automated analysis process and outcomes
- Video shows multi-image blending technique
  - Two different datasets (X-axis)
  - Dynamic range compression at different frequency cutoffs (Y-axis)

Kovesi, P.D., Holden, E.J., Wong, J.C. 2014, 'Interactive multi-image blending for visualization and interpretation', Computers & Geosciences, v. 72, pp 147-155.

Lessons learnt

- Good algorithms were not enough (requiring time consuming false positive removal)
- A workflow was designed
  - Understanding the nature of the data; and how end-users use/interpret the data
  - Initial exploration of data to identify regions where automated processing may not be reliable
  - To give control of the process to the user as much as possible (e.g. parameter selection)
  - Interactive visualisation to communicate intermediate results & **confidence/uncertainty** in the results, and to incorporate user input
  - Allowing post-processing (removing false positives)
- Deployment was through commercialisation - as a plug-in to industry standard software and through a marketing and licensing agreement with the software vendor

My Thoughts on Future of Data Science

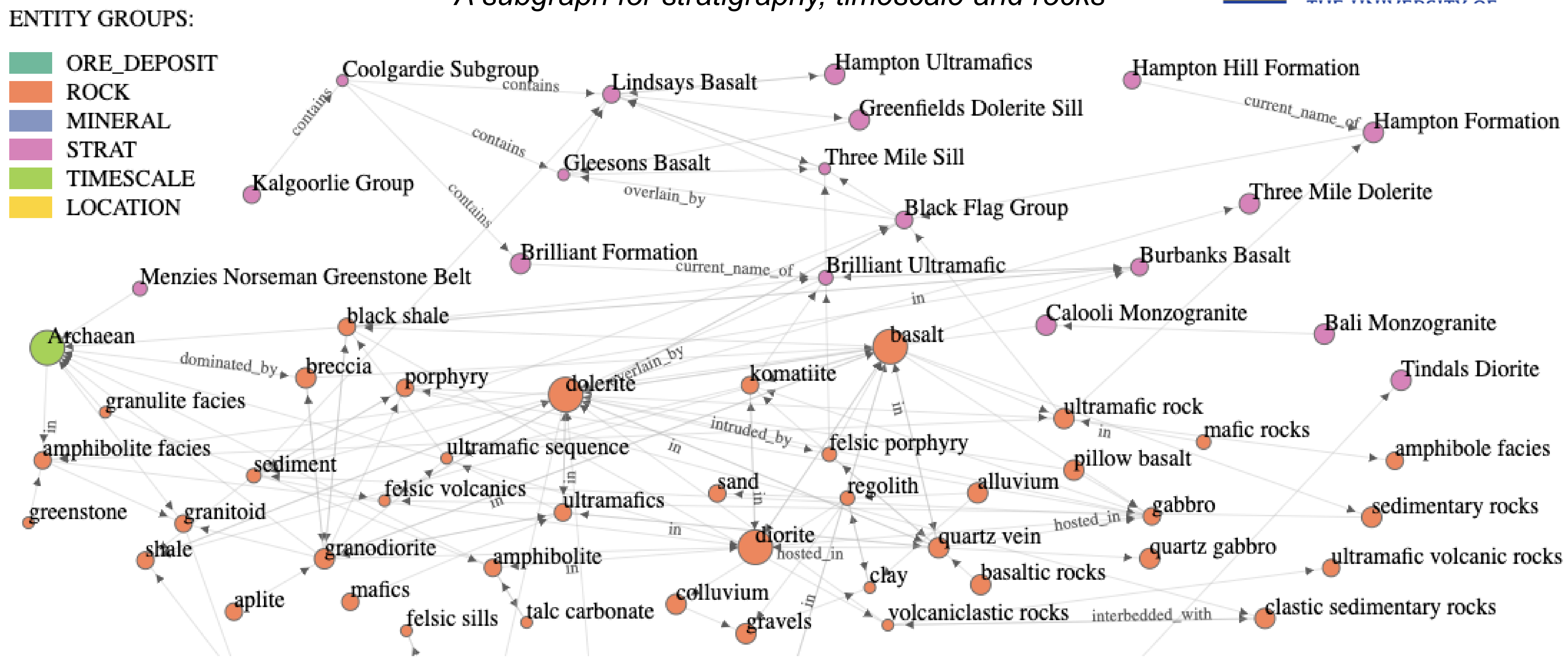
- Increasingly seeking Data-Driven Decisions for those with significant social, financial and environmental implications
  - Machine Augmented vs Automated data analysis
- Practical data science needs to be entwined with domain/industry specific understanding of data and analytical processes
- Progressing from information discovery (pattern recognition) to knowledge discovery (insights)
  - Interlinking information/knowledge from different types of data (structured, unstructured & semi-structured; multi-modal data, etc)
- Ethical and socially acceptable use of data and automation became of paramount importance for industry and society broadly

Discovering insights from text

Majigsuren Enkhsaikhan

- **Knowledge Retrieval**
- PhD research by *Majigsuren Enkhsaikhan* (CSSE, UWA, Supervisors - Wei Liu, Eun-Jung Holden, Paul Duuring)
- *Challenge*: Building a robust conceptual model for mineral exploration framework is challenging
- *Approach*: Extract, store and retrieve the relations between locations, geological era, stratigraphy, rocks, minerals and ore/deposits using **Knowledge Graph** (Graph based database)
  - The first step towards interlinking unstructured data with structured data & AI question and answer tool

A subgraph for stratigraphy, timescale and rocks



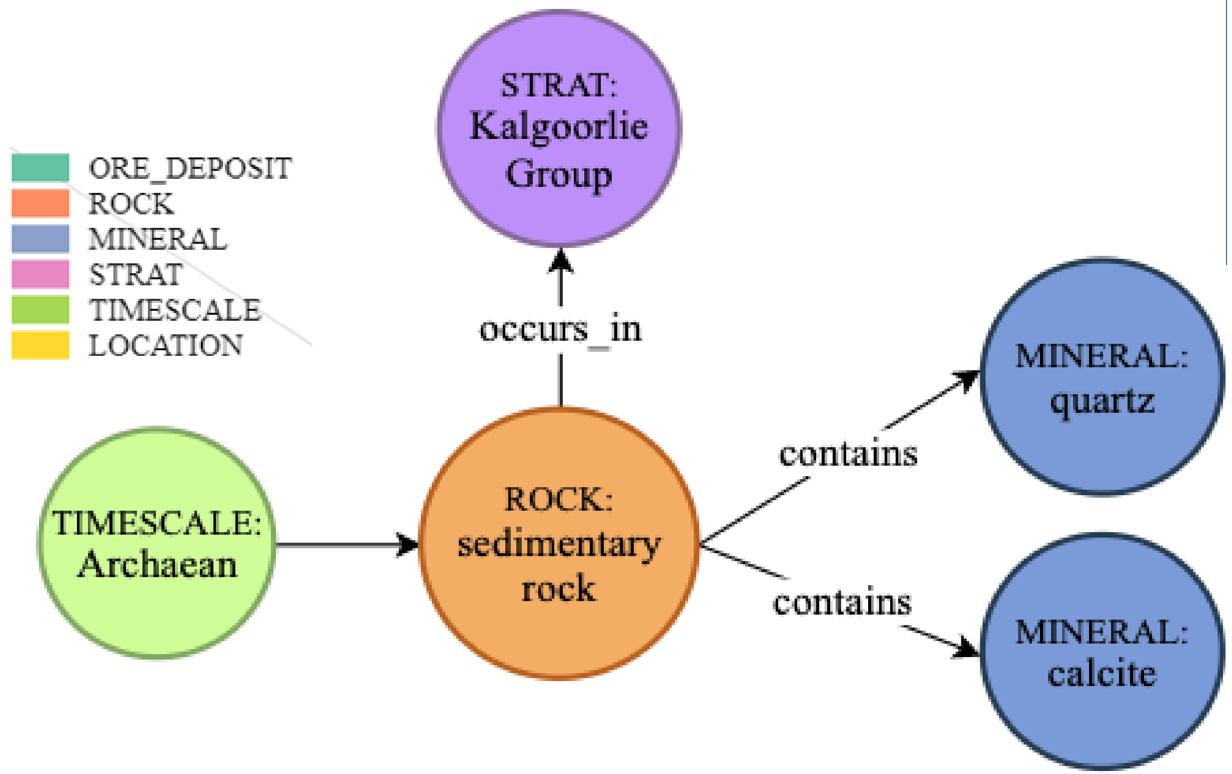
Stratigraphy: The relations include **Kalgoorlie Group** contains **Coolgardie Subgroup**, which contains **Lindsays Basalt**, **Gleesons Basalt** and **Brilliant Formation**; **Black Flag Group** overlain by **Gleesons Basalt**; **Brilliant Formation** current name of **Brilliant Ultramafics**.

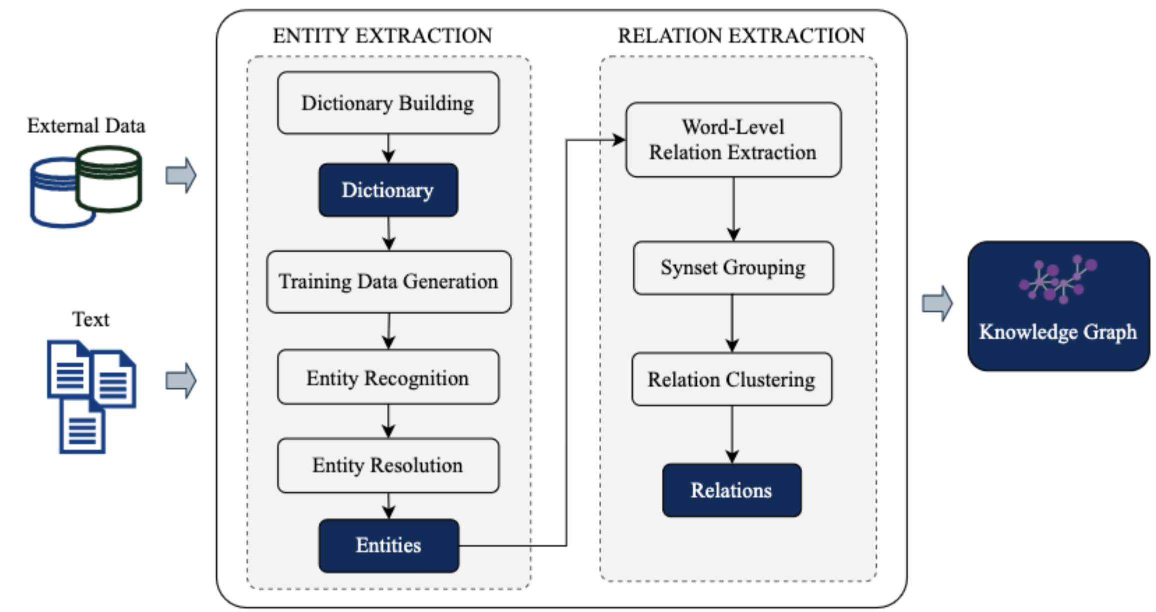
INTERLINKING OF Australian Stratigraphic Unit Database (ASUD) for stratigraphy hierarchy

A. Archaean sedimentary rocks occurred within the Kalgoorlie Group.
Most sedimentary rocks contain either quartz or calcite.

B. Archaean ——> sedimentary rocks
sedimentary rocks —occurred within—> Kalgoorlie Group
sedimentary rocks —contain—> quartz
sedimentary rocks —contain—> calcite

C.





← Knowledge Graph

Refs

- Sivarajah et al., 2014, Visual Saliency and Potential Field Data Enhancements: Where is your attention drawn?, Special section: Interpretation and integration of gravity and magnetic data, Interpretation, November, pp. SJ9-SJ21.
- Sivarajah et al. 2014, Quantifying Target Spotting Performances with Complex Geoscientific Imagery Using ERP P300 Responses, International Journal of Human-Computer Studies 72, 3, pp. 275-283.
- Drew, T., Võ, M.L.H, and Wolfe, J.M., The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers, Psychological Science, September 2013 24: 1848-1853
- Wedge, D., Holden, E. J., Dentith, M. & Spadaccini, "Fast and objective detection and analysis of structures in downhole images", 2017, Journal of Applied Geophysics. 144, p. 157-172
- Holden, E. J., Liu, W., Horrocks, T., Wang, R., Wedge, D., Duuring, P. & Beardsmore, T., 2019, GeoDocA – Fast analysis of geological content in mineral exploration reports: A text mining approach, Ore Geology Reviews, 111
- Vasuki Y., Holden, E.J., Kovesi, K., Micklethwaite, S., 2017, An interactive image segmentation method for lithological boundary detection: A rapid mapping tool for geologists, Computers & Geosciences, v.100, pp 27-40
- Majigsuren Enkhsaikhan, Eun-Jung Holden, Paul Duuring and Wei Liu. "Understanding ore-forming conditions using machine reading of text". In: Ore Geology Reviews. 2021, p. 104200
- Majigsuren Enkhsaikhan, Wei Liu, Eun-Jung Holden and Paul Duuring. (2021). "Auto-labelling entities in low-resource text: a geological case study". In: Knowledge and Information Systems, pp. 1–21

CITS5553 Projects & Groups, 2022

Projects

1. Predicting Conductive Ear using WBA & WBT (Audiology, UWA)
2. Anomaly based Conductive Ear Detection using WBA (Audiology, UWA)
3. Energy Usage Prediction (Data Analysis Australia)
4. Predicting Beach Wave Condition using offshore buoy data (Oceans Institute, ICRAR, UWA)
5. Analyse and Visualise the Effects of Public Health Restrictions on Perth Crime (WA POLICE)
6. Rapid, accurate and precise identification of mineral grains (Portable Spectral Services)
7. Machine learning estimation of in-patient metrics (Department of Health)
8. On the applicability of NLP methods in patient triage (Department of Health)

P1: Predicting Conductive Ear using Wide Band Absorbance (WBA) & Wid Band Tympanometry (WBT)

Clients:

- Robyn Choi (robyn.choi@uwa.edu.au); Chris Gonzalez (chris.gonzalez@uwa.edu.au)

Question: Can machine learning be applied to existing wideband absorbance (WBA) and wideband tympanometry (WBT) data to improve understanding and interpretation by providing an automatic diagnosis of ears with conductive conditions?

Project Objective

1. To detect and understand which frequency regions in WBA and WBT results are most useful for detecting conductive conditions in children;

2. To investigate whether machine learning tools is more accurate at detecting ears with conductive condition compared to receiver operating characteristics analysis.

Team

22374535 Gunadi	Mr	Dennis
21520699 Ika	Mr	David Leonardo
21124413 Jiang	Mr	Andy
23220451 Kurniady	Mr	Edward
22614179 Tan	Mr	Andrew
21918545 Xie	Miss	Jiaxin

P2: Anomaly based Conductive Ear Detection using Wide Band Absorbance (WBA)

Clients:

- Robyn Choi (robyn.choi@uwa.edu.au); Chris Gonzalez (chris.gonzalez@uwa.edu.au)

Question: Can data for children with normal wideband absorbance (WBA) be characterised using an existing dataset, and using these two combined datasets, can anomalous signals for conductive ears be detected?

Project Objective

1. To characterise normal WBA data using an existing dataset, and if possible, generate false normal data that aligns the probability distribution of the existing dataset of normal ear;

2. To automatically detect anomalous or out of distribution behaviours or signals for conductive ear detection.

Team

22755019 Le	Miss	Thi Phuong Thao
23266049 Li	Miss	Jiaze
23246811 Nandasena	Mr	Athaudage Charana Yasasvin
22856585 Pananchickal Asok	Mr	Govind
22507957 Tan	Mr	Benjamin Shong Hao
21321139 Wilyman	Mr	Callum Peter

P3: Energy Usage Prediction

Clients

Brett Andrijich brett@daa.com.au **(Data Analysis Australia)**

Background: Prediction of the days that have high energy consumption is important for energy suppliers and management of power grids.

Requirements: A model that for each day predicts what the peak energy usage for that day will be (based on half hour intervals) and provides a recommendation whether it is likely to be one of the top four peaks energy usage days. Variables to be included in the model TBD by group but can include rainfall, temperature, humidity, calendar dates etc

Team

21719308 Bui	Mr	Danny
23035606 Harris	Mr	Isaac Syaukani
22492426 Joyner	Mr	Samuel James
23136706 Nizarudeen	Mr	Naufal
23057103 Sharma	Mrs	Vandana
22889911 Zhang	Mr	Zheyu

P4: Predicting Beach Wave Condition using offshore buoy data

THE UNIVERSITY OF WESTERN AUSTRALIA

Supervisors

Kevin Vinsen (kevin.vinsen@icrar.org) and Jeff Hansen (jeff.hansen@uwa.edu.au)

Background

Each wave buoy records it motion in both the vertical and horizontal directions which is then processed (using spectral analysis techniques) to produce a statistical representation of the wave field typically over 30-60 minutes. The most detailed statistical representation is the so called 2D spectrum which shows the wave energy as a function of frequency (period) and direction.

Objective

This project aims to use wave buoy data collected offshore of Western Australia to predict the wave conditions at a Perth beach. This project is to using buoys located in Cape Naturaliste, Mandurah, and Rottnest; it will aim to predict the 2D spectrum at Cottesloe beach, where a buoy will provide the ground truth

Team

21963986 Jain	Mr	Varun
23292749 Liu	Mr	Jiachuan
22643009 Mashaire	Mrs	Thanks
23006014 Spencer-Laitt	Mr	Patrick Jacob
21955556 Van Den Broek	Mr	Jacob Cees
22483045 Zhuang	Ms	Jia Ni

P5: Analyse and Visualise the Effects of Public Health Restrictions on Perth Crime

THE UNIVERSITY OF WESTERN AUSTRALIA

Client

Shih Ching Fu, Projects Coordinator, WA Police Force

shihching.fu@police.wa.gov.au

Background

Anecdotally the patterns of crime in WA were affected by the public health restrictions instituted during the first few waves of the COVID-19 pandemic. Crimes against property and drug seizures appeared to decrease during lockdowns whilst other crime types remained unchanged or even increased in frequency. Furthermore, a significant proportion of WA Police Force resources were diverted towards enforcing public health policies, potentially reducing capabilities focussed solely on crime.

Objective:

Identify changes in crime patterns in WA that are possibly caused by the restrictions and policing resourcing changes imposed during the COVID-19 pandemic.

Team

22256457 Byala	Ms	Megan Nora
23193697 Chen	Ms	Haixin
23372936 Kamboj	Mr	Ankit
20757557 Kochhar	Mr	Bhuvnesh
23264879 Ng	Mr	Yat Wah
21484971 Smith	Mr	David Phillip Timothy

P6: Rapid, accurate and precise identification of mineral grains

THE UNIVERSITY OF WESTERN AUSTRALIA

Client

Dr Nigel Brand, Consulting Geochemist, Portable Spectral Services

nwbrand@portaspecs.com

Background

Portable Spectral Services is in the process of automating the micro- Xray fluorenes technology (µXRF) to increase through-put from 40 samples per day to 200 samples per day. Micro-XRF is a scanning technology that rapidly maps the distribution of elements in a sample at the micro scale. In this case the samples are heavy mineral concentrates (HMC's). HMC's are high valuable samples that are being used to identify mineral systems.

Project Objectives

- Evaluate the application of data science in identifying mineral grains using the micro-XRF technology.
- Determine a data science workflow that can be validated by a spectral geologist.
- Elucidate, which elements are critical and which elements are obsolete in the identification of mineral phases.

Team

22166122 Alhasan	Mr	Ali Yaseen H
22504484 Fung	Mr	Kevin Che-Hang
23225847 George	Ms	Vidhya
22253888 Hooper	Mr	Austin Douglas
22541459 Zhang	Mr	Hanlin

P7: Machine learning estimation of in-patient metrics

Client

Shiv Meka, Department of Health

shivakarsh.meka@health.wa.gov.au

Background

In Australian hospitals, unscheduled emergency admissions affect the planning and scheduling of resources - hospital beds, staff, and clinical equipment (X-rays, CT/MRI, etc.). A major problem that affects Western Australian hospitals is patient flow. Synonymous with the term "flow", patient flow relates to the difference between the number of patients admitted and discharged in a given time and under the assumption that discharged patients have been subject to "proper" care during their hospital stay.

Project Objectives

- The main aim is to predict four important in-patient metrics: admission risk, a patient's length of stay, mortality risk, and 30-day avoidable hospital readmission (AHR) risk.

Team

22001577 Abbasi	Mr	Shahzad Anwar
23087175 Holt	Mr	Samuel George
23185196 Lai	Mr	Hok Him
23161711 Nguyen	Ms	Thuy Linh
22262605 Varsani	Mr	Vikesh Shamji

P8: On the applicability of NLP methods in patient triage

Client

Shiv Meka, Department of Health

shivakarsh.meka@health.wa.gov.au

Background

Patient triaging is a process where patients visiting the emergency department are prioritised based on the presenting symptom. Triaging is performed by a trained triaging nurse who, depending on a subject's symptoms, assigns a score. Australian hospitals conform to what is called Emergency Severity Index (ESI) which is a score between 1-5 (1 for the most severe).

Project Objectives

- A machine learning model that returns a triage score based on patients' data (esp. free text).
- The model clusters related conditions/records.
- Visualisation to improve usability

Team

22483359 Zhou	Mr	Nathan
23191175 Sigar	Mr	Jeremiah Junior
22095334 Zhou	Ms	Jiaxin
23033292 Muthyam	Mr	Akhil
21984315 Bagley	Mr	Kai Adam

Support

Mentors

- Dr Tom Horrocks (tom.horrocks@uwa.edu.au)
- Dr Daniel Wedge (Daniel.wedge@uwa.edu.au)
- Will be available to provide support for the choice of data science techniques for all projects
  – one hr session for each group for the proposal development; and another session for the progress

My consultation time: Mondays 2-4pm (F2F or via Teams)

Conflict resolution: please see me (the unit coordinator) as the first of call

Data Science – *A broad discipline that uses diverse analytical and visualisation approaches and addresses challenges in producing real world products*

For Interpreters

- Data sparsity and uncertainty
- Integrated interpretation of data collected from different modalities with different sampling resolutions
- Presupposition/bias plays an important role

For Data Scientists

- Difficult to model and validate the outcomes
- Data integration
  - Patterns look different in different modality data
  - Different resolutions
  - Missing data

- Have a clear value proposition (Transformational)
- Algorithm processes are understandable as much as possible (Interpretable, transparent)
- Geologically feasible solutions; seamless integration to interpreters' workflow (Actionable)

Smart routing of automatic number plate recognition (ANPR) equipped police vehicles to crime scenes.

Client

Shih Ching Fu, Projects Coordinator, WA Police Force

shihching.fu@police.wa.gov.au

Background

When a major (criminal) incident occurs, often several police vehicles are dispatched to the scene simultaneously. Many of these vehicles may be equipped with automatic number plate recognition (ANPR) capability which affords the opportunity to log incidental intel as these police vehicles proceed to the scene. This intel consists of the time and location of other vehicles that are passed enroute to the scene. These data may prove useful for example if a suspect’s vehicle is identified along one of the routes departing the scene.

Objective:

Develop a vehicle routing algorithm that considers the location of a major incident, local traffic patterns, and the disposition of nearby ANPR (un)equipped WA police vehicles; to coordinate and optimise the capture of valuable ANPR data while still ensuring timely arrival at the scene of the incident.

Team