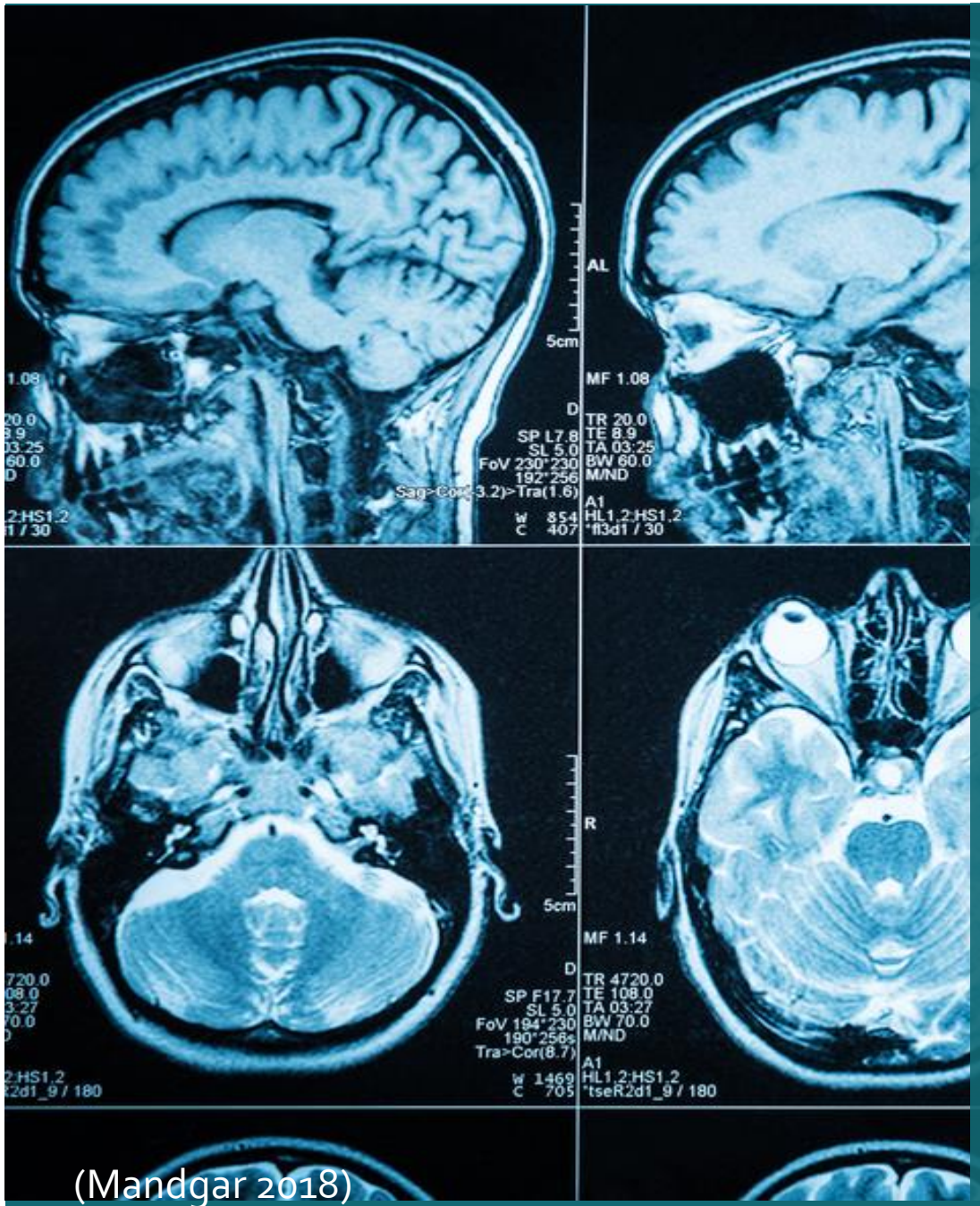
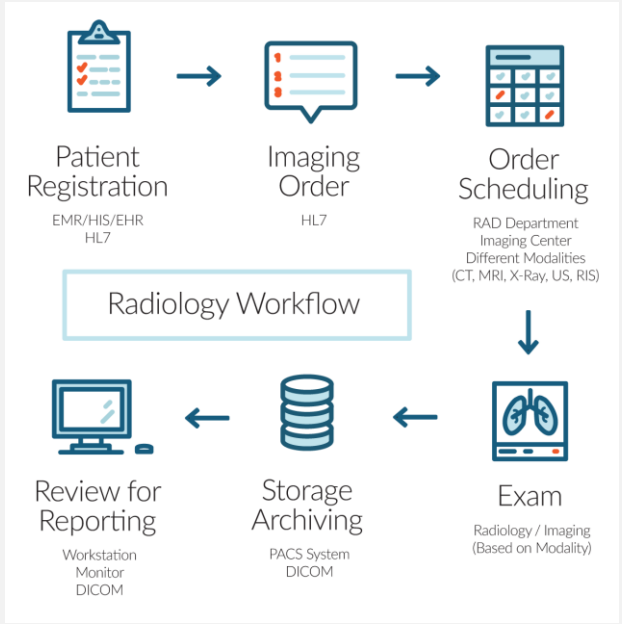


Data Science Project: Radiology Machine Learning

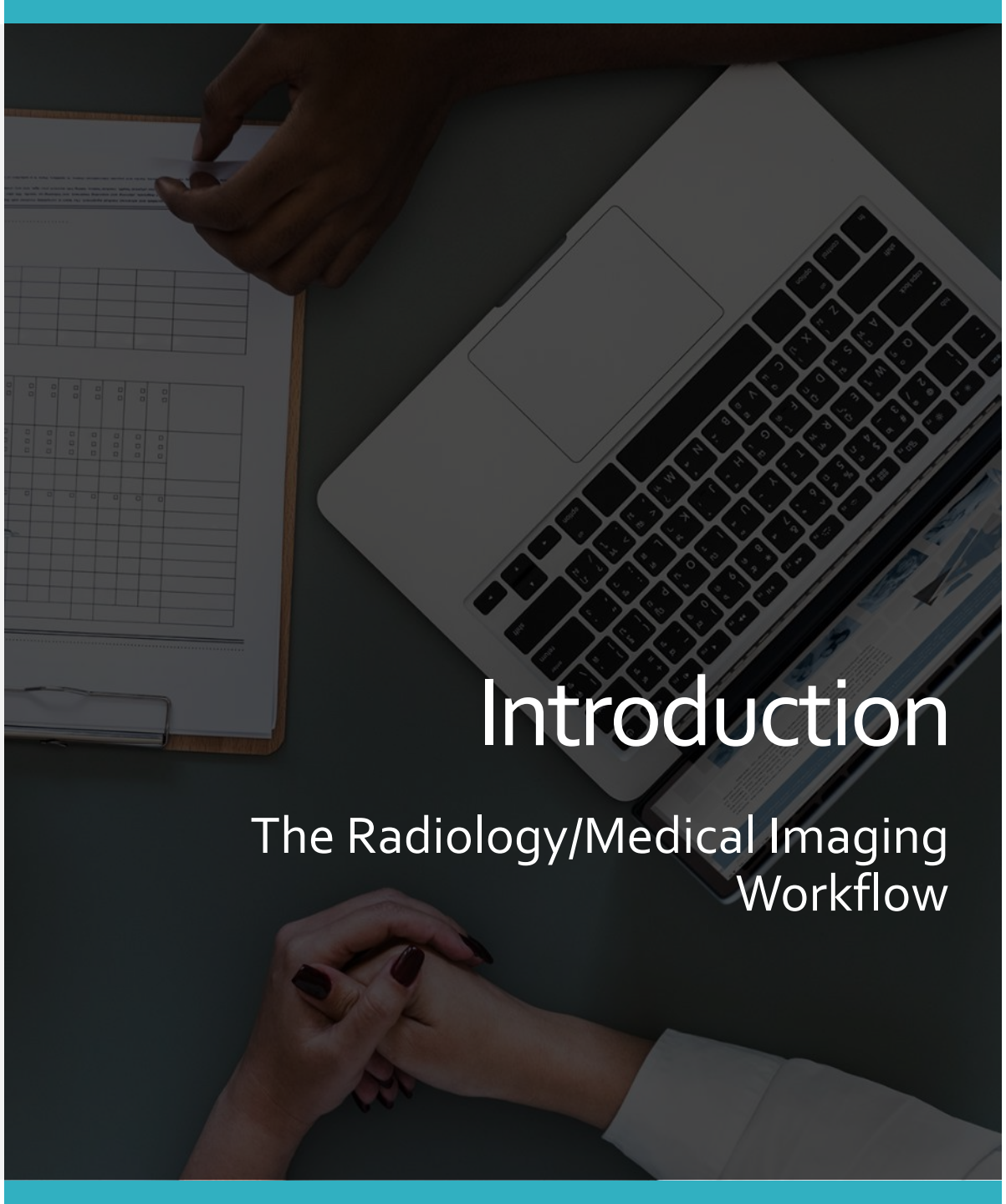
Quality Improvement Project

(Mandgar 2018)



(Datar 2016)

- Increasing workload with an aging population
- Shift from Private health to Public Health
- Radiology is a specialty focused on pattern recognition
- Medical imaging images are popular for machine learning due to their structure



Introduction

The Radiology/Medical Imaging Workflow

The Problem

Increasing Workload

Radiology Department is getting busier

Limited Resources

Equipment is old and not much money is spent on updating

Staff Inexperience

Lack of training and relevant soft skills

Errors Spread Faster

Highly interconnected systems

Incorrect Scans and Reports

Labelling and Classification Errors

The Problem

(Radiopaedia.org 2019)

Correct Laterality

Study Description: Left Foot

Incorrect Laterality

Study Description: Left Foot

(MeVis 2019)

Potential Outcomes

- Coded as Right Foot
- Report Refers to Right Foot
- Teams assume report is correct and may apply treatment to Right foot incorrectly
- Extreme example is incorrect surgery!

Solution

How to address this without more humans.

Realtime

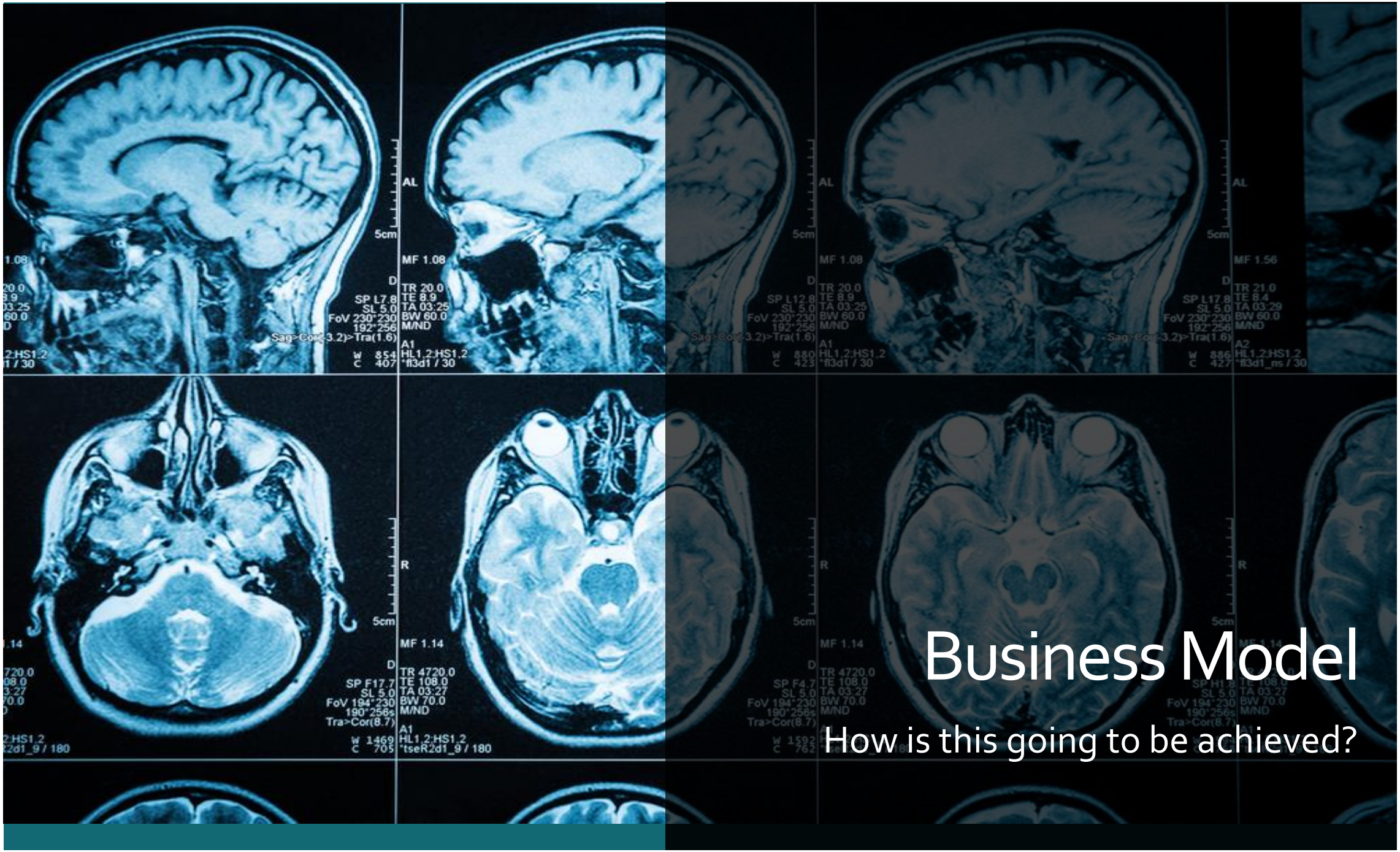
Access data fast

Alerts

Notify by the exception

Proactive

Catch issues before they spread



Business Model

How is this going to be achieved?

The Data Science Team

Data Developer

(Application Developer)

Data Creative

(Imaging Informaticist)

Data Business Person

Head of Radiology

Data Researcher

(Quality Manager)

(Buntime 2017)

Business Model

Data Collection

- Collate medical imaging to central archive
- Creation of the data pipeline (front end and backend)
- Data Creative/Developer

Engineering

- Images extracted from database, Loaded into Python using Keras , NumPy and Matplotlib
- Data Creative/Developer

Governance

- Patient privacy and confidentiality
- Cyber Security
- Laws and Legislation
- Data Business Person / Researcher

Wrangling

- Images augmented, scaled and saved to JPEG
- Train and Validation and Test sets
- Data Creative/Developer

Analysis

- Classification model trained
- Prediction match label?
- Data Creative/Researcher

Presentation

- Live Dashboard
- Email Alerts
- Presentation Summary
- Data Business Person

Challenges

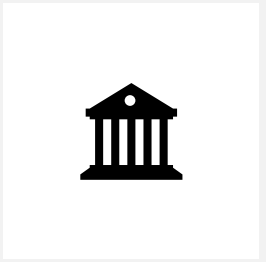
Some of the main things to consider

Confidentiality

Patient data is sensitive

Liability

Mistakes can cost financially, reputation and more



Regulation

Need to comply with local , state and federal data laws

Benefits

Lots to gain

Reduce Risk

Catch errors earlier

Improve Care

Reduction in patient harm

Efficiency gain

Frees up human's time

Data "The 4 V's"

Volume

- How much data?
  - Each DICOM image is about 10-12 MB
  - JPEGs are much smaller (400KB)
  - 400 images in training
  - 60 per day
  - 30 GB per Year

Velocity

- The frequency the data needs to be processed?
  - As soon as possible (real-time)
  - Early anomalies can be found the less the impact on the business
  - Limited by access to production server

Veracity

- How trustworthy is the data?
  - Mostly trustworthy due to patient checks
  - Depends of experience and competence of radiographer
  - Both feet under single foot code
  - Toes under foot code

Variety

- What are the kinds of data?
  - DICOM images
  - Structured data with a metadata model
  - JPEG images

Tools and Software

Python

- Python (Python 2019) will be used for the wrangling and analysis
  - Open source and well supported
  - Use main packages such as Keras , Matplotlib, Numpy and Scikit Learn
  - Compatible with all other products in the project
  - Well supported with strong community

Database (SQL)

- Postgres SQL (PostgresSQL 2019) will be used as the database
  - Open source with widely available support
  - Fast performance
  - Easy to learn and use
  - Well supported with strong community

PACS (DCM4CHEE)

- DCM4CHEE (DCM4CHEE 2019) will be used to process the data
  - Open source with excellent standards and API for access
  - Able to host locally
  - Robust and is scalable
  - Well supported with strong community

Microsoft Power-Bi

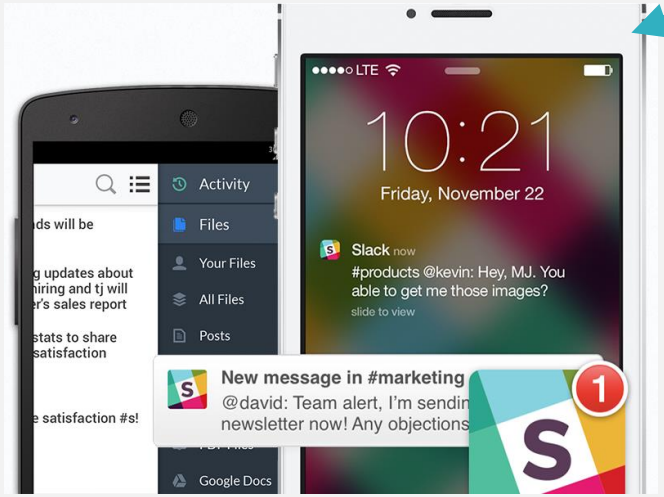
- Microsoft Power BI (Microsoft 2019) will be used for the presentation and analysis
  - Free to use (desktop version)
  - Easily integrates with data sources such as Postgres and excel
  - Can use with python scripting
  - Create live dashboards
  - Can also get commercial support

Workflow

QEPACS2

DICOMweb™

WAPACS



- Split the data into Train, Validation and Test Sets
- Augmentation of images to improve recognition (Scale, Rotate)
- Transfer Learning
- Learning Rate
- Activation Functions
- View performance
  - Accuracy
  - Loss

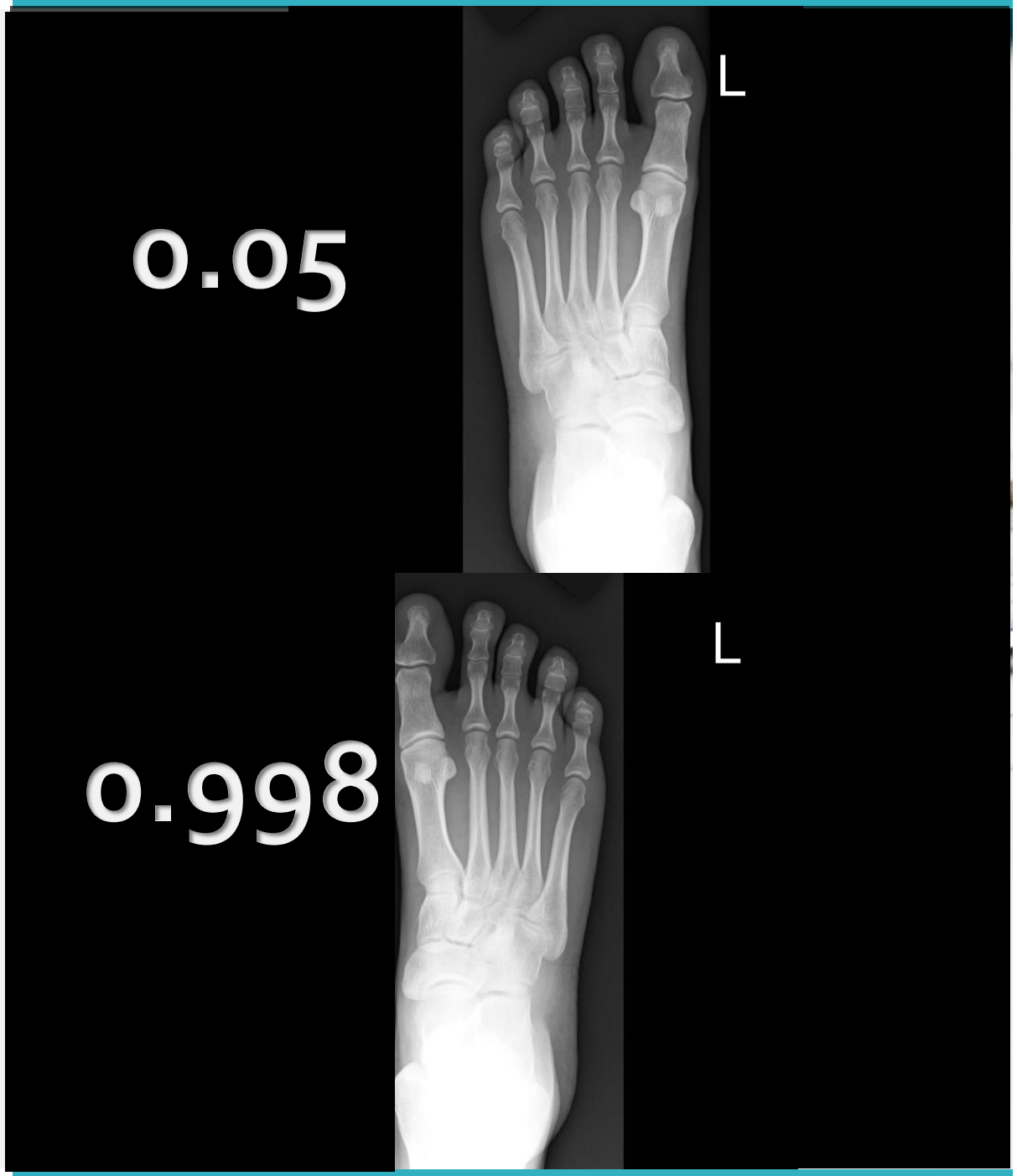
- Model is trained to perform image classification (supervised learning)

Wrangling

Getting the data ready

- Model is now trained to perform image classifications
- Images are pulled into Python via wrangling pipeline and analyzed
- Model scores images on a scale of 0 to 1 on whether it is a left or right foot
- Compares this to the image metadata
  - If no match this triggers the alert workflow to get further investigation
- Results saved to the Database

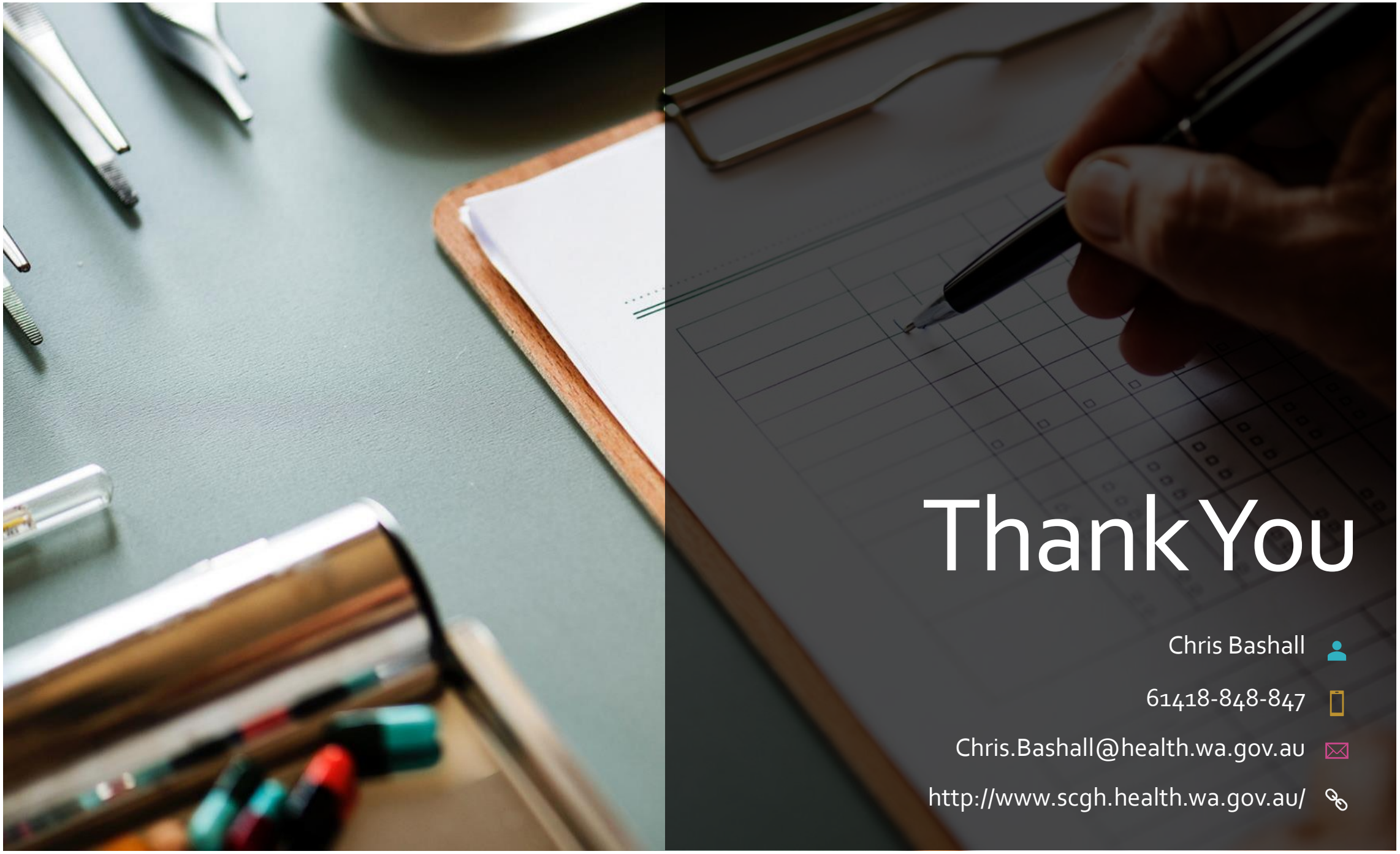
- Power-Bi (Microsoft 2019) is used to extract the metadata from the database over time and present performance including
  - Precision
  - Recall
  - Accuracy
  - Time Saved



Summary

Outputs of the project

- Small, low risk introduction to a data science project in a production environment
- Help identify some of the challenges and benefits that may arise from data science projects
- Show the potential and perhaps wet the appetite?



Thank You

Chris Bashall

61418-848-847

Chris.Bashall@health.wa.gov.au

http://www.scgh.health.wa.gov.au/