



# **Analytics in Medical and Health - from Data to Application**

Dr Yu Yong Poh

<https://www.richieeyptutorialpage.com/>

R Conference 2023

R User Group Malaysia (MyRUG)

## **Disclaimer:**

The views and opinions expressed in this presentation are solely those of the presenter and do not necessarily reflect the views, opinions, or policies of any company, organization, or community group. This presentation is intended **for informational and educational purposes only**. Any references to specific companies, products, or services are for illustrative purposes and should not be considered as endorsements.

The presenter **does not have any official affiliation or representation** with any company or community group mentioned during this talk. All information and content presented here are based on personal knowledge and experiences.

Please consult with appropriate professionals and organizations for specific advice and recommendations related to the topics discussed in this presentation.

# What will be covered?

- Introduction to Health Analytics
- Historical Context and Evolution
- Types of Data in Healthcare
- Data Management Challenges
- Key Analytics and Techniques and Tools
- Impact of Analytics on Patient Care
- Future Trend and Innovations
- Challenges and Considerations

# 1: Introduction to Health Analytics

- A discipline that combines data, information technology, statistical analysis, and quantitative methods to derive insights for decision-making and management in healthcare settings
- Data quality - enabling a systematic analysis of healthcare big data
- Big data analytics - facilitating personalized healthcare

## 2: Historical Context and Evolution

- Big Data - offering insights and improving patient outcomes
- Big Data Challenges - Data fragmentation and Ownership
- Example: Next Generation Sequencing (NGS) - sequencing genomes and other 'omics' data; 'genomical' data to reach 2–40 Exabytes by 2025
- Machine Learning and Artificial Intelligence -  
Eg: Electronic Health Records

# 3: Types of Data in Healthcare

- **Electronic Health Records (EHRs):** demographics, medical history, allergies, laboratory test results, and billing information.
- **Imaging Data:** X-rays, MRIs, and CT scans.
- **Genomic Data:** genomic sequencing generates large volumes of data that can inform personalized treatment plans and predictive health models.
- **Patient-Generated Data:** data from wearable devices and patient-reported outcomes - providing real-time insights into patient health status and behaviors
- **Pharmacy Data:** prescriptions, medication adherence information, and pharmacy refill records - medication management and patient compliance.
- **Insurance and Billing Data:** This includes claims and payment data, which can be analyzed for health service utilization, cost of care, and health economic studies.
- **Social Determinants of Health (SDOH):** socioeconomic status, education, and environmental factors, which significantly influence health outcomes.

# 4: Data Management Challenges

- **Fragmentation:** Big Data in healthcare can be fragmented across different systems and providers - difficult to integrate and utilize for comprehensive patient care.
- **Volume and Complexity:** significant challenges in storage, analysis, and interpretation.
- **Privacy and Data Ownership:** data ownership, sharing, and privacy while aligning with regulations like PDPA, GDPR and HIPAA.

# 5: Key Analytical Techniques and Tools

- **Descriptive Analytics:** Use historical data to create reports and dashboards that provide insight into what has happened in the past in a healthcare setting.
- **Predictive Analytics:** Employ statistical models and machine learning algorithms to predict future events based on historical data, such as patient readmission rates or disease outbreaks.
- **Prescriptive Analytics:** Utilize advanced analytics and mathematical models to provide recommendations for clinical decision-making, resource allocation, and treatment options.
- **Machine Learning and AI:** pattern recognition, diagnostic aid, and even in predictive models for patient treatment plans
- **Data Visualization Tools:** Tableau, Qlik, and PowerBI - visualizing complex datasets in an interpretable and actionable format



# 6a: Use Case: TB Advisor - AI Chatbot

TB Advisor ▾



You

what are the TB symptoms



TB Advisor

Tuberculosis (TB) symptoms can vary depending on whether the TB is latent or active.

1. **Latent TB:** In latent TB, the bacteria remain in the body in an inactive state. They cause no symptoms and are not contagious, but they can become active.
2. **Active TB:** This condition can affect the lungs (pulmonary TB) or other parts of the body (extrapulmonary TB). Common symptoms of active TB, particularly pulmonary TB, include:
  - Persistent cough that lasts for three weeks or longer
  - Coughing up blood or sputum (mucus from deep inside the lungs)
  - Chest pain, or pain with breathing or coughing
  - Unintentional weight loss
  - Fatigue
  - Fever
  - Night sweats
  - Chills
  - Loss of appetite

For extrapulmonary TB, symptoms depend on the organs involved. For example, TB of the spine may cause back pain, and TB in the kidneys might cause blood in the urine.

It's important to consult a healthcare provider for a proper diagnosis and treatment if you or someone you know is experiencing these symptoms. TB is a serious condition but can often be cured with appropriate treatment.



<https://chat.openai.com/g/g-HvEOeScUo-tb-advisor>

# 6b: Use Case: TB Advisor - AI Chatbot



<https://chat.openai.com/g/g-HvEOeScUo-tb-advisor>

 You

write codes to classify if a x-ray lung image is health or has tb

 TB Advisor

Developing a machine learning model to classify lung X-rays as healthy or indicative of tuberculosis (TB) is a complex task that requires substantial data, computational resources, and expertise in medical imaging and machine learning. However, I can provide a high-level overview of how one might approach this task using Python and common machine learning libraries.

## Steps to Build a TB Classifier

1. **Data Collection:** You need a dataset of lung X-ray images labeled as 'healthy' or 'TB'. Public datasets like the [NIH Chest X-ray Dataset](#) can be used as a starting point.
2. **Data Preprocessing:** Preprocess images for consistent size, resolution, and contrast. Data augmentation (like rotations, flips, etc.) can be used to increase the dataset size.
3. **Model Selection:** Choose a model architecture. Convolutional Neural Networks (CNNs) are typically used for image classification tasks. Pre-trained models like ResNet, VGG, or Inception can be fine-tuned for this specific task.
4. **Training:** Split the dataset into training, validation, and test sets. Train the model on the training set, using the validation set to tune hyperparameters.
5. **Evaluation:** Evaluate the model's performance on the test set using metrics like accuracy, precision, recall, and F1-score.
6. **Deployment:** For practical use, the model needs to be deployed in an environment where clinicians can use it to assist in diagnosing TB from X-ray images.

## Sample Code Structure (Pseudocode)

```
python
import numpy as np
import pandas as pd
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Conv2D, Flatten, MaxPooling2D
from tensorflow.keras.preprocessing.image import ImageDataGenerator

# Load and preprocess data
# Assuming X_train, X_val, X_test, y_train, y_val, y_test are prepared

# Data augmentation
train_datagen = ImageDataGenerator(rescale=1./255, rotation_range=10, width_shift_range=0.1, height_shift_range=0.1)
val_test_datagen = ImageDataGenerator(rescale=1./255)
```

## 6c: Other Use Cases

- **Project InnerEye by Microsoft:** This project employs computer vision and machine learning to differentiate between tumors and healthy anatomy using 3D radiological images
- Many More ...

# 7: Impacts of Analytics on Patient Care

- **Personalized Medicine:** Analytics enables a tailored approach to treatment, considering a patient's genetic makeup, lifestyle, and other personal factors.
- **Predictive Health:** Use of predictive models can identify patients at high risk for certain conditions, allowing for early intervention and preventive care.
- **Operational Efficiency:** Analytics can streamline hospital operations, leading to reduced wait times and improved patient experience.
- **Outcome Improvement:** By analyzing trends and outcomes, healthcare providers can improve the quality of care and patient satisfaction.
- **Cost Reduction:** Analytics can identify cost-saving opportunities without compromising the quality of care, such as by predicting which patients are likely to be readmitted and taking preventative measures.

## 8: Future Trends and Innovation

- **Integration of Genomic and Clinical Records:** providing a more comprehensive view of patient health that can lead to more personalized care.
- **Interoperability of Electronic Health Records (EHRs):** The development of standards like Fast Healthcare Interoperability Resources (FHIR) is aiding in the seamless exchange of healthcare information across different systems.
- **Blockchain in Healthcare:** securely and privately transfer and store healthcare data, ensuring patient control over their information.
- **EHRs as Diagnostic Aids:** transforming EHRs from mere repositories of patient information into diagnostic tools

# 9: Challenges and Considerations

- **Interoperability of EHRs:** different systems and providers often use incompatible formats
- **Outdated Technology:** outdated technology can pose security risks and inefficiencies
- **AI Transparency:** The "black box" nature of AI and machine learning models is a concern
- **Algorithmic Bias:** AI models can perpetuate existing biases found in the data they are trained on

