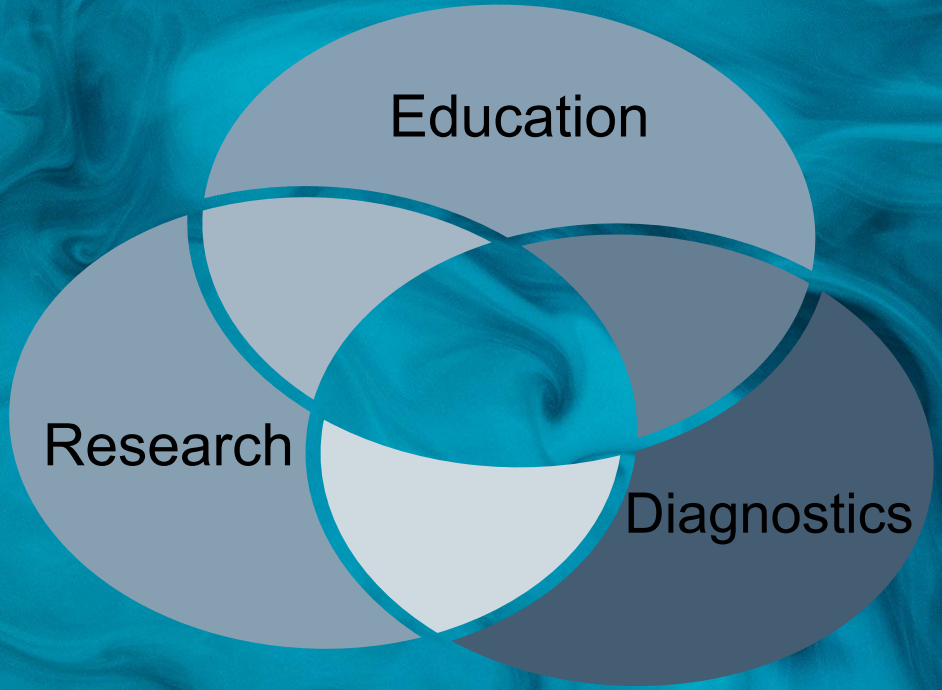
Capstone Project Presentation

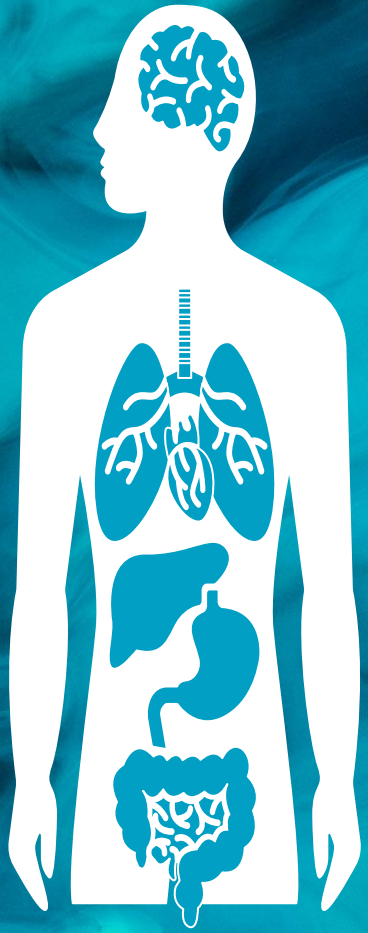
Tresha
10 June 2021

READ

Research Education and Diagnostics



What does your gut say?



- Colorectal cancer one of leading causes of cancer death in the U.S
- Early diagnosis prevents surgery and low quality of life
- More check-ups and regular scopes increase workload of pathologists

Research EducationAand Diagnostics

Applications

- Interactive website as support tool
- Improve diagnosis accuracy
- Training/Education purposes
- Research Purposes

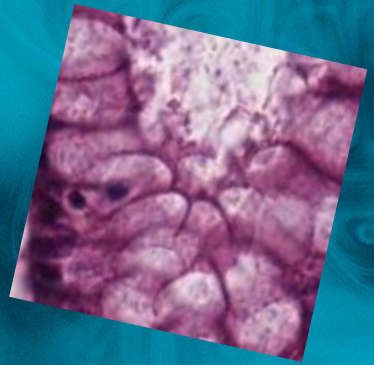
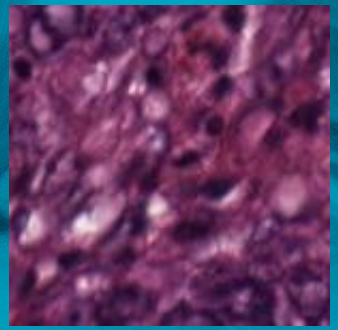
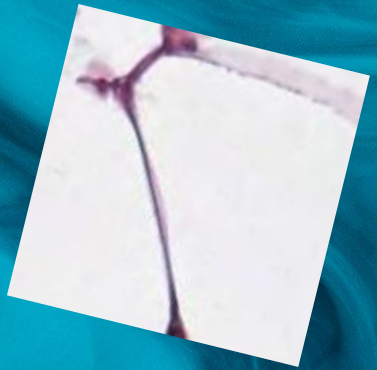
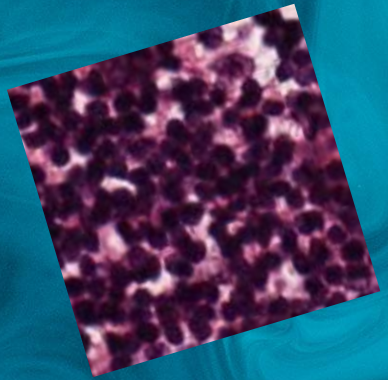
Research EducationAand Diagnostics

Data Sets Used

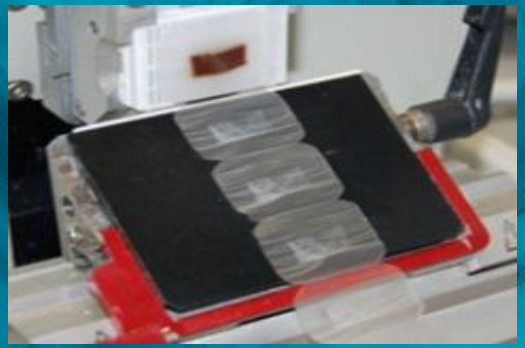
Capstone (Diagnosis): Colorectal Cancer Image Data

Extension (Treatment): Clinical Evidence Text Data

Colorectal Cancer Image Dataset :

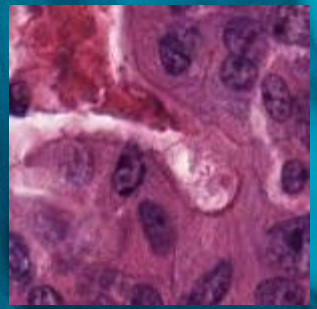
Background Information



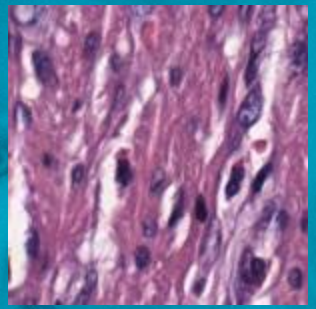
- Patient samples (tissue) are fixed in wax
- Thin sections of wax with tissue are sliced
- Sliced sections are stained and examined

Colorectal Cancer Dataset

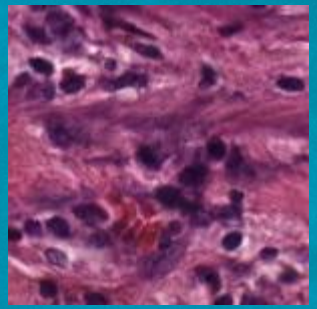
1)Tumour



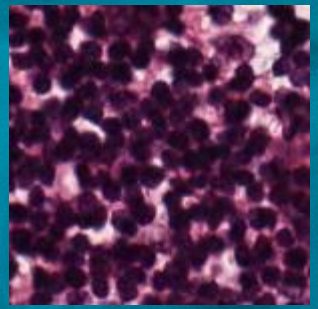
2)Stroma



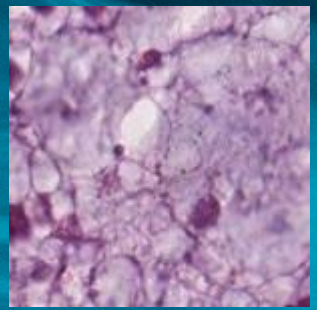
3)Complex



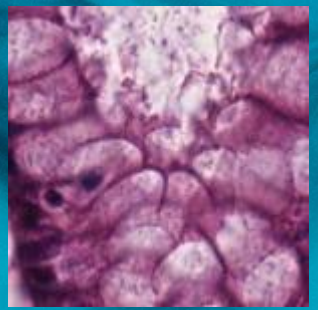
4)Lympho



5)Debris



6)Mucosa

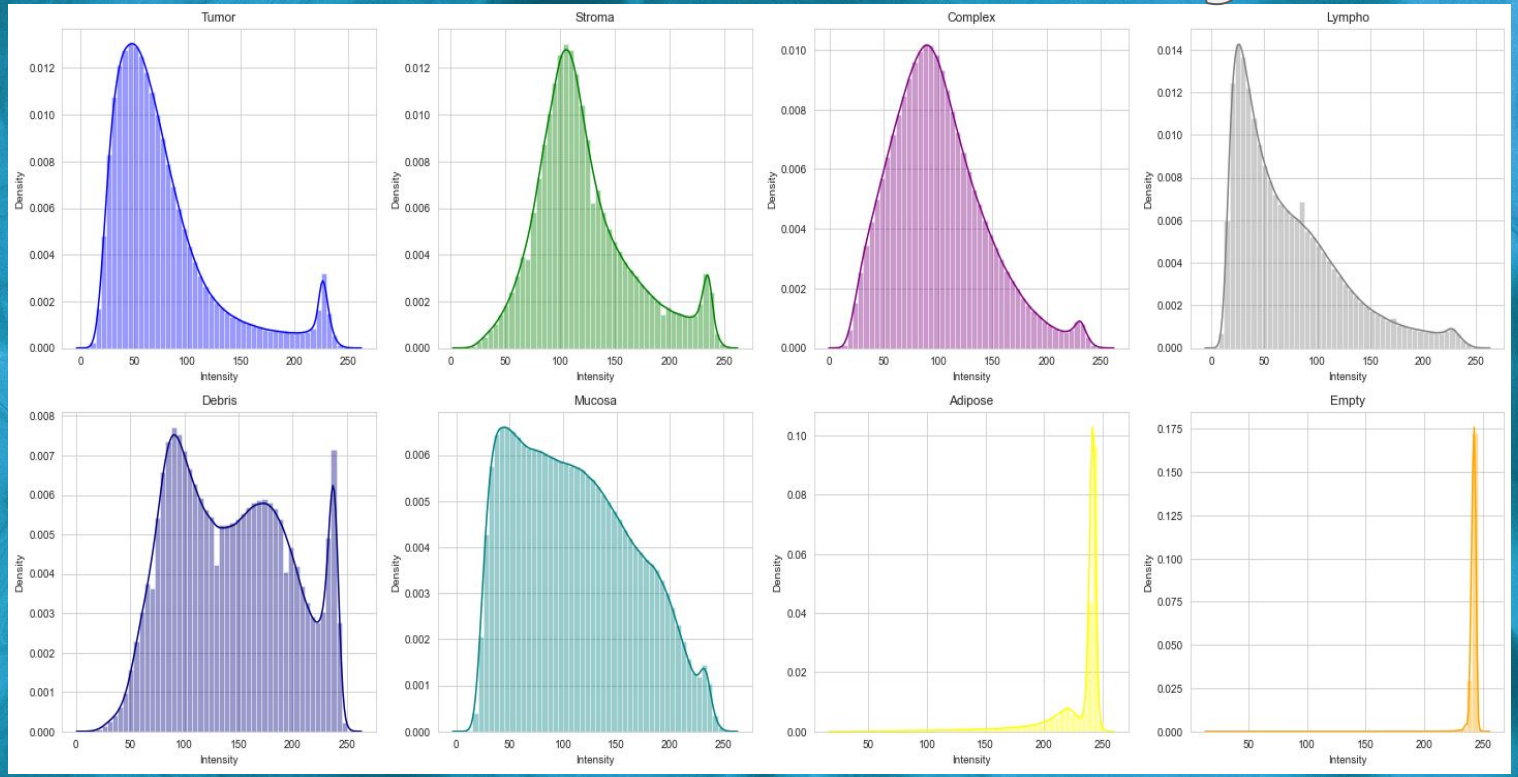


7)Adipose

8)Empty

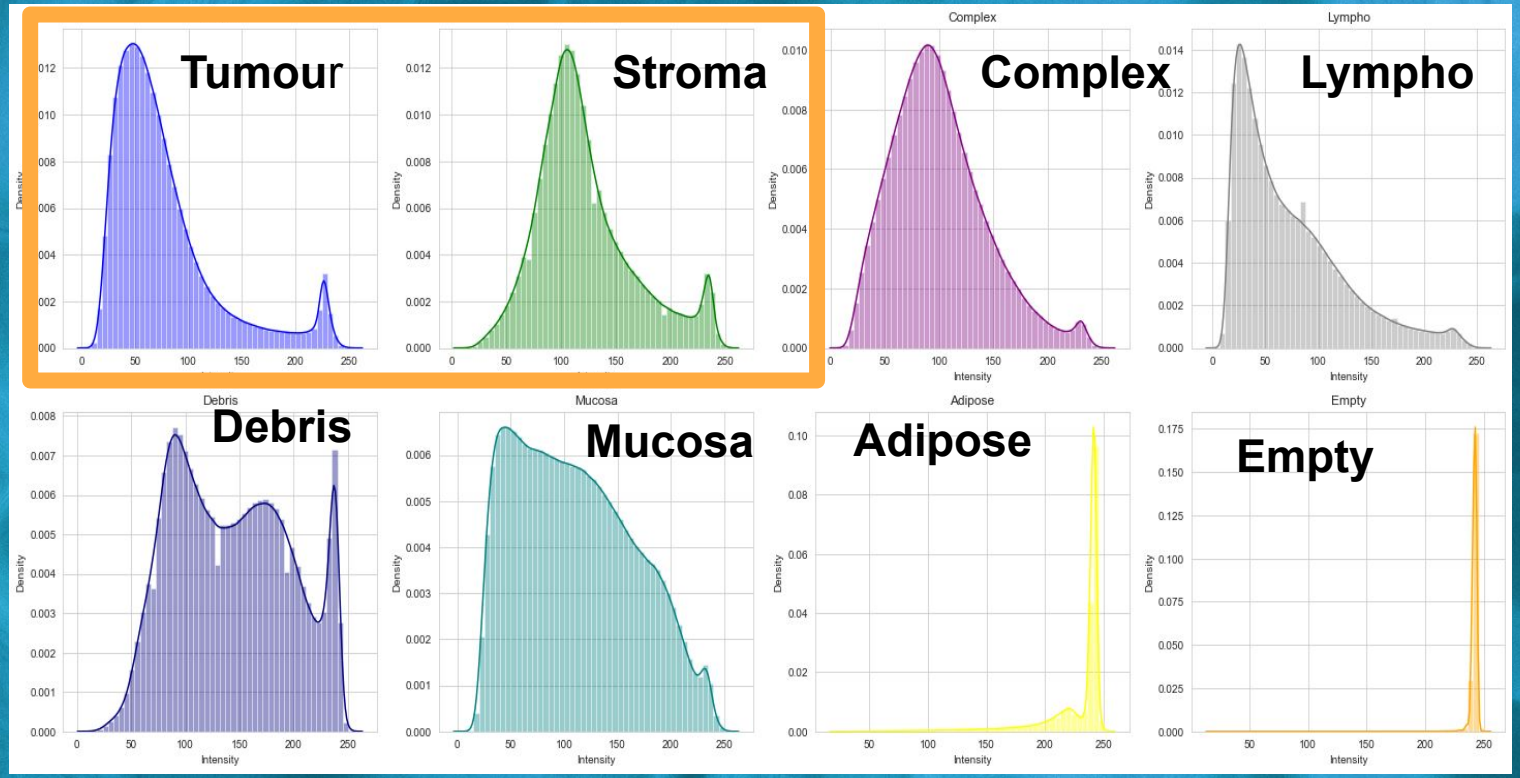
8 types of tissues from colorectal cancer image data

Pixel Data Analysis



DENSITY

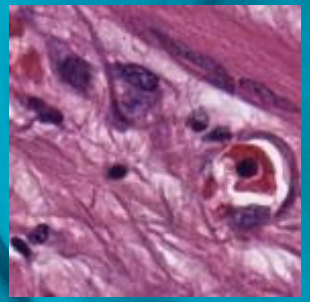
INTENSITY



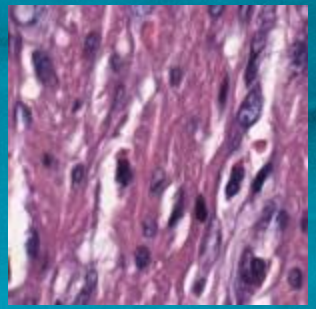
DENSITY

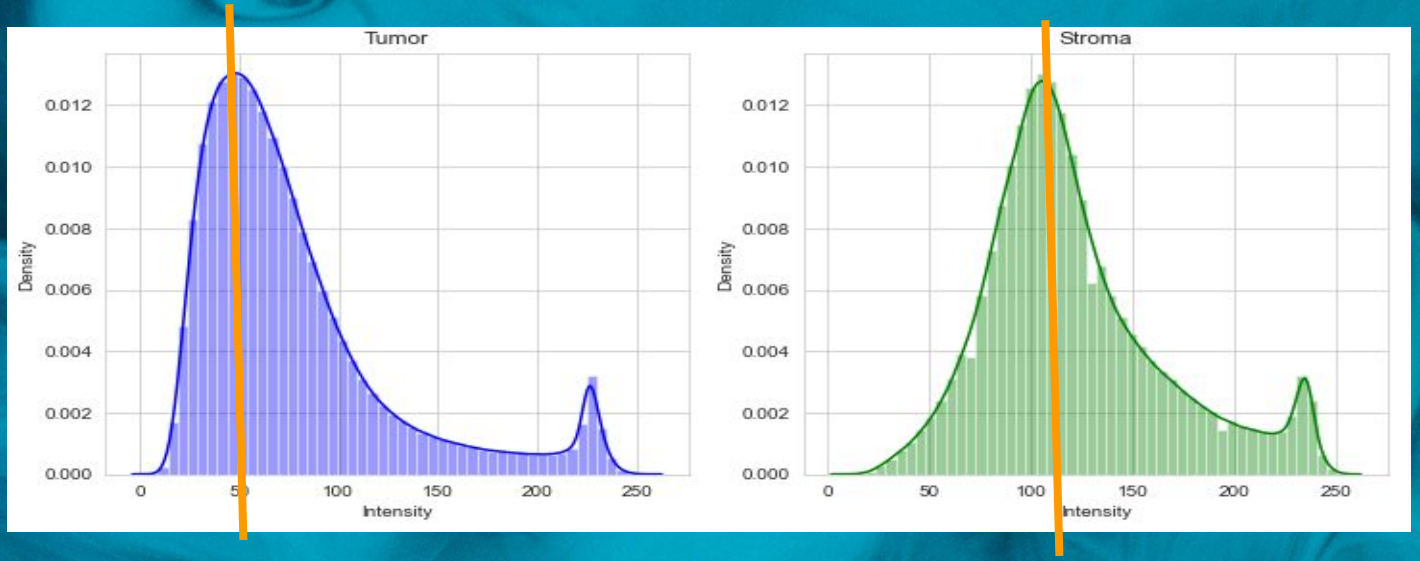
INTENSITY

1)Tumour



2)Stroma





Tumour VS Stroma

Tumour VS Stroma Classification Model

- More accurate diagnosis by differentiating between tumour and stroma tissues
- Stroma-rich tissues have  overall survival
- Stroma-rich tissues have  disease-free state
- Stroma:Tumour ratio

Modelling for colorectal cancer image pixel data

Models	Accuracy Score	Baseline Score
Random Forests	0.947	0.5
Convulated Neural Network	0.960	0.5
Support Vector Machines	0.925	0.5

Pixel values Data VS Images Data

Pixel Data	Image Data
0.96 accuracy score with CNN	0.71 score using CNN
Lesser memory space since numbers	More memory space with images stored
More data can be added for various types of cancers	Lesser data results in limited use of website created for classification

Deploy classification model on website in future

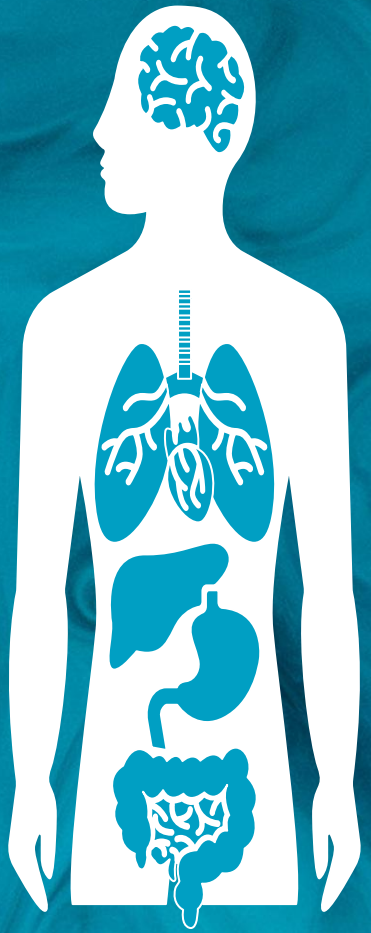
- Interactive Software developed in future
- Upload Image
- Annotate the area of interest
- Software converts image to pixel values
- Model classifies into Tumour or Stroma
- Software provide Tumour: Stroma ratio

Capstone Extension: Clinical Evidence Text data

Clinical Evidence Text Data

- Pathologists classify mutations from patient samples' gene sequencing data
- Review Clinical Literature
- Try to personalize treatments if it's available
- Reviewing and classifying is time-consuming

- This process can be automated to safe time
- Natural Language Processing for classification
- Platform to identify the class of mutation in a more standardised manner
- Provide personalised treatment if available
- Researchers can use the platform to find novel treatments



Research EducationAand Diagnostics

Clinical Evidence Text Data is classified into 9 classes of mutations

Mutation Class	
1	[A111P, A1843P, A209T, A290T, A4419S, A636P, C...
2	[A1374V, A2034V, A2425T, A859_L883delinsV, ACP...
3	[A41P, A41T, D2512G, D2512Y, F1888V, G264S, I1...
4	[3' Deletion, A1022E, A120S, A1685S, A598T, C1...
5	[A1066V, A272V, D32Y, E116K, E31K, E501G, E541...
6	[Amplification, C528S, D603G, E172K, E501K, E5...
7	[422_605trunc, A11_G12insGA, A1459P, A146T, A1...
8	[E40N, G311D, HMGA2-RAD51B Fusion, K700R, S492...
9	[R132G, R132H, R140Q, R172S, R625C, R625H, R62...

- **Naiive Bayes perform better**
- **F1 score used due for imbalance classes**
- **Use Flask for prediction on website**

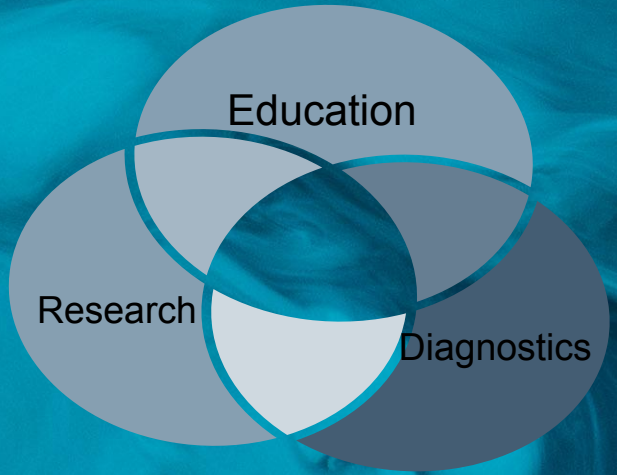
		Accuracy	Precision	Baseline Score for Largest Class
1	Naiive Bayes	**0.600**	**0.620**	0.379
2	KNN	0.515	0.510	0.379
3	SVC	0.602	0.575	0.379
4	Random Forest	0.610	0.600	0.379
5	Logistic Regression	0.572	0.544	0.379

Executive Summary:

- Machine learning models to classify clinical text and image data
- Support tools for diagnosis and education purposes
- Spearhead research to find novel treatments for personalised medicine

Modelling:

- Pixel Data takes less memory
- Classification of tumour and stroma using Convulated Neural Networks gives an accuracy score of 0.96
- Naiive Bayes performs better for NLP data for imbalanced classes



READ

Research Education and Diagnostics

Future Work:

- Image classification interactive platform to be productionised
- Clinical Evidence Text Model to be improved by increasing colorectal cancer dataset

Research EducationAand Diagnostics

Website :

R.E.A.D

Research Education and Diagnostics

You've come to the right place: a platform to share images and text data to get answers to your questions and all the discussions in between. Take a look at each channel and start your own engaging conversation today!

Thank you for your attention!