SUMMER INTERNSHIP REPORT

AT

(15 April to 28 June 2024)

A Report

By

Mr. Swastik Mitra

PG/23/120

Under the guidance of: Dr. Preetha GS

POST GRADUATE DIPLOMA IN HOSPITAL AND HEALTH MANAGEMENT 2023-2025

International Institute of Health Management Research New Delhi

ACKNOWLEDGEMENTS

I would like to express our sincere gratitude to the following individuals and organizations for their valuable contributions to this report:

I am extremely grateful to Dr. Vidhur Mahajan, CEO of CARPL AI, for giving me the opportunity to intern under his esteemed guidance. His mentorship, invaluable insights, and constant encouragement have greatly enriched my learning experience.

Mr. Rinku Yadav, for their expertise and guidance throughout the research process.

Dr. Amit Kumar, for providing access to essential resources and for their assistance in analysis and interpretations.

I am also deeply thankful to Dr. Preetha GS, my mentor at the International Institute of Health Management Research. His continuous support, guidance, and constructive feedback have been instrumental in shaping my academic and professional development. Her expertise and dedication have provided me with a strong foundation for my future career.

Additionally, I would like to extend my heartfelt thanks to the faculty and staff of the International Institute of Health Management Research, New Delhi for their unwavering support and for providing a conducive learning environment. Their commitment to excellence has significantly contributed to my growth and success.

I am grateful for the support and encouragement we received from everyone involved in this project.

(Completion of Summer Internship from CARPL.ai Pvt. Ltd.)

The certificate is awarded to

Name: Swastik Mitra

In recognition of having successfully completed his internship in the department of

Title: HCP onboarding and demand generation

and has successfully completed his Project on

AI-Driven Radiology: Performance Insights and Clinical Implications

Date: 28/06/2024

Organisation: CARPL.ai Pvt. Ltd.

He comes across as a committed, sincere & diligent person who has a strong drive & zeal for learning

We wish him all the best for future endeavors

Organization Supervisor

Head-HR/Department Head

CARPL.AI PVT. LTD.
New Delhi

FEEDBACK FORM
(Organization Supervisor)

Name of the Student: SWASTIK MITRA

Summer Internship Institution: CARPL.ai. Pvt.Ltd

Area of Summer Internship: Demand Generation, Database Management, HCP onboarding & Marketing

Attendance: 100%

Objectives met:
1. Learnt about strategies of demand generation.
2. Learnt about AI ecosystem
3. Learnt automation tools - Apollo.io
4. Prepared database for different strategies - ICP Reachouts
5. Conferences & CARPL Awareness Programms

Deliverables:
1. Database - ICP, Conferences (SIIM), Carpl awareness program
2. ROI analysis of Modalities in Australia
3. Analysis of AI performance at clinical site.

Strengths:
Database Management, Workload handling, Modality Research. Good at handling Softwares such as Apollo.io, Power BI, Excel & powerpoint

Suggestions for Improvement:
—

Signature of the Officer-in-Charge (Internship)

Date: 28 June 2024
Place: Saket, Delhi
CARPL.ai Pvt. Ltd.

FEEDBACK FORM

(IIHMR MENTOR)

Name of the Student: SWASTIK MITRA

Summer Internship Institution: CARPL. ai. Pvt. Ltd.

Area of Summer Internship: Demand Generation, Database Management, HCP onboarding & Marketing

Attendance: 100%.

Objectives met:
1. Learnt about Strategies & functioning of corporate businesses.
2. Learnt about AI ecosystem & Sales process in AI driven health organization
3. AI testing & monitoring model
4. Research, Data analysis using Softwares, Report formation along with regular workings on feedbacks

Deliverables:
1. Research Report, ppt and poster on AI driven Rad: Performance Insight & clinical implication
2. Delivered presentation on the research conducted withing Institution & organization

Strengths: Meticulous, hard working.

Suggestions for Improvement:
Create more models for AI testing & evaluation to get a better exposure of quality while using AI

Signature of the Officer-in-Charge (Internship)

Date: 5 July 2024
Place: Dwarka, Delhi
IIHMR-D

Certificate of Approval

The Summer Internship Project of titled **AI-Driven Radiology: Performance Insights and Clinical Implications** at **CARPL AI** is hereby approved as a certified study in management carried out and presented in a manner satisfactorily to warrant its acceptance as a prerequisite for the award of **Post Graduate Diploma in Health and Hospital Management** for which it has been submitted.It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed, or conclusion drawn therein but approve the report only for the purpose it is submitted.

Dr. Preetha GS
Professor and Dean Research
IIHMR, Delhi

TABLE OF CONTENTS

ABBREVATIONS

AI	Artificial Intelligence
CARPL	CARING Analytics Platform
HCP	Healthcare Provider
RIS	Radiology Information System
PACS	Picture Archival Communication System
FDA	Food and Drug Administration
EC	European Conformity
HAS	Health Sciences Authority
ANVISA	Brazilian Health Regulatory Agency
TGA	Therapeutic Goods Administration
VC	Venture Capitalist
KPI	Key Performance Indicator
API	Application Programming Interface
PHI	Personal Health Information
IUID	Item unique identification
MSK	Musculoskeletal
CTA	Call To Action
ICP	Ideal Customer Profile
POC	Proof of Concept

NDA	Non-Disclosure Agreement
USG	Ultrasound
MRI	Magnetic Resonance Imaging
AUC	Area Under Curve
ROC	Receiver Operating Curve
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
CRM	Customer Relation Manager
ASX	Australian Securities Exchange
PAS	Platform as Service
SAS	Software as Service
ROI	Return on Investment
GT	Ground Truth

OBSERVATIONAL LEARNING

SECTION-1: INTRODUCTION

1.1 ABOUT CARPL:

CARPL.ai is the world's first testing and deployment platform for medical imaging AI applications, which connects healthcare providers to third party AI applications, helping improve access, affordability, and quality of medical care.

It bridges the gap between healthcare providers and AI developers by serving as a gatekeeper that seamlessly connects both sides of the ecosystem. In essence, it is a single interface to access AI algorithms, validate and test them, and subsequently embed them into radiology workflows.

It is used by some of the world's leading healthcare providers, AI researchers, industry teams and startups. It is built with the single goal of making it easy for clinicians to get access to advanced analytics tools.

1.2 VISION

CARPL's vision is to be the back-end platform behind all medical imaging AI deployment globally by becoming the single interface for AI deployment at healthcare providers, and the go-to-market strategy of choice for AI developers.

1.3 FEATURES

- Easy to Access: There is a fragmented ecosystem out there if any HCP wanted to deploy any radio diagnostic AI solution how they can contact AI developers providing similar solutions individualistically, but CARPL platform have 100+ AI solutions of different AI developer on a simple single user- friendly platform
- Easy to Assess: CARPL is the first company to provide the feature of Validation and Testing & Monitoring of the AI solutions for the HCPs so they can analyze the efficiency of the AI model over their own Patient data.

- Easy to Integrate: If the HCPs starts deploying AI solutions related to radiodiagnosis by multiple AI developers for different use cases they would require different RIS-PACS integrations but for accessing multiple AI solutions via CARPL requires single RIS-PACS setup.
- Reliability: All the AI solutions onboarded on CARPL's platform are regulatory compliant including FDA(US), CE(Europe), HSA(Singapore), ANVISA(Brazil) and TGA (AUS).

SECTION 2: MODES OF DATA COLLECTION

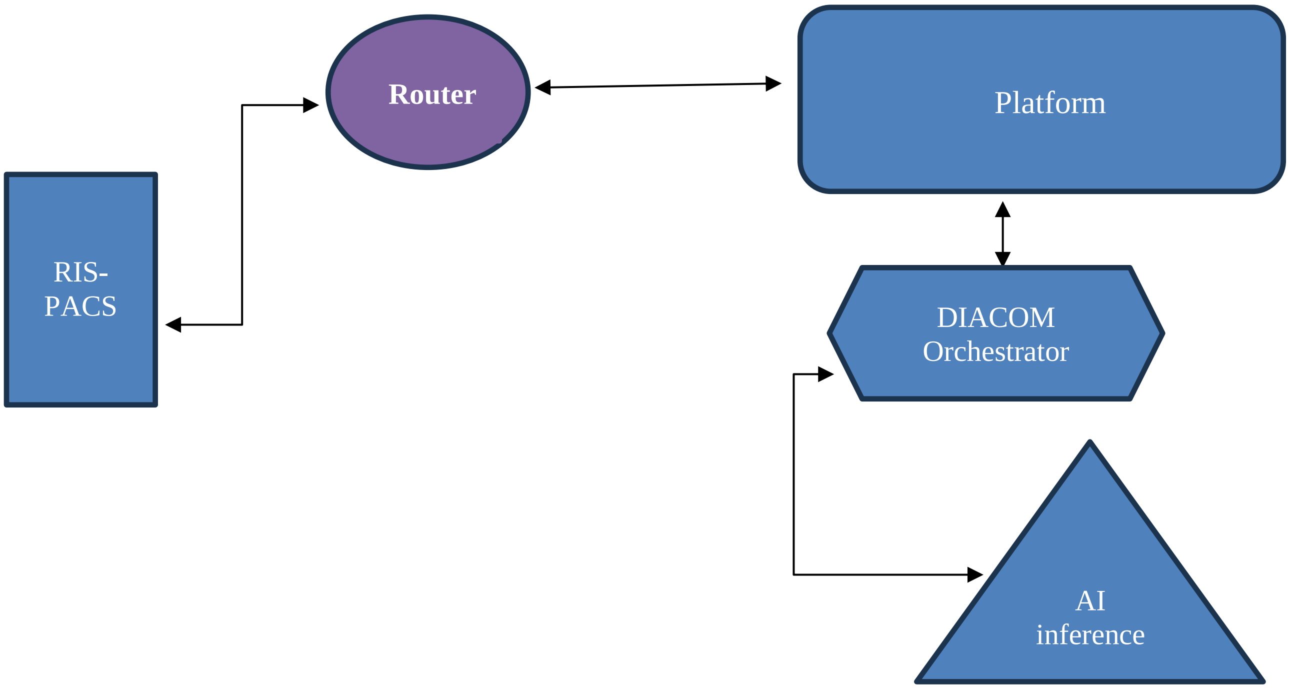
- Observation
- Document Analysis
- Unstructured Interviews

SECTION 3: DEPARTMENT'S OBSERVATION AND FINDINGS

➢ **Operations Department:**

- Headed by the Chief Operating Officer (COO)
- Daily operations, equipment management, finance, travel, Human Resource management and infra level operations all are being managed by the Operations Department.
- Human resource management is playing a very crucial role as CARPL is in the transitional phase from a startup to a funded organization (VC) and is continuously striving to get constant sustainable growth so requires competent human resources. Active interviews, referral programs, actively hiring for positions like – Interns, Product Development Manager, Business Development Manager and Analysts.
- Weekly full team meetings where every department shares the weekly progress and happenings with the other departments and discussions on different crucial topics takes place.
- Automated HR analytics for tracking KPI's, attendance, leave applications, salary etc. using HR one Software.
- Meeting schedules are generally managed by Google calendar
- According to the job requirement all the IT essentials are being provided to the employees like laptops, Email ids, Monitors, printers etc.

- **Technology Department:**
  - Headed by the Chief Technology Officer (CTO)
  - The IT team handles the platform level operations which can be understandable by this diagram:



  - In the above figure, multiple stages of the image process, transfer and AI operation over the image and result return are shown.

  - The IT team is divided into engineers working on these specific stages that are:
    1. Engineers are working for the smooth integration of the RIS and PACS of the clients with the CARPL platform so that the clients won't face any type of issues related to breakdown times, uploading studies on platform, data breaching and security concerns. Teams ensures the images which are being uploaded with any means like the DIACOM push, CARPL console or the API should be seamless for the clients end along with the monitoring correct DIACOM tags linked with the appropriate image for successful uploading and AI inferencing.

2. Engineers working on the router the primary work of the router is to capture the uploaded dataset from the client's PACS and anonymization of the personal health information (PHI) of the patients and filter the DICOM tags according to the images pushed by the PACS.

Router Anonymize the data by providing.

Study IUID (MSK XRAY scans dataset)

Series IUID (Shoulder XRAY Scans within Study scans)

SOP IUID (Each view of the XRAY scan)

Engineers manages the Frontend and Backend of the router for successfully executing these operations.

3. Engineers handle backend and frontend of the platform for the clients to use the AI model enabled platform efficiently without any bugs and lags. At the backend dataset management, regular updates, AI inferencing, platform optimization and bug fixations are the prime focus whereas frontend engineers optimize the user experience on the platform with regular feature additions.
4. The same journey of the scans repeats after the AI inferencing but in the reverse direction from platform to the PACS and this time router deanonymizes the scans and along with AI outputs transfer to the PACS server.

AI output can be of-

a) DIACOM GSPS
b) DIACOM structure report
c) Secondary report
d) Encapsulated

- **Marketing Department:**
  - In CARPL marketing department completely focuses on the MCI approach:
    1. Marketing: Marketing Mix model for fulfilling the objectives
    2. Communication: Use of the right channel for efficient exchange of ideas
    3. Information: Complete Information about the product and the market in which product will be launched

  - Target Audience:

    Geographically:

    i. US
    ii. Australia
    iii. Singapore
    iv. Brazil

    Organizations:

    i. Health Systems
    ii. Hospitals
    iii. Radiology Groups & Societies
    iv. Universities

    People:

    i. Chief Executive Officer
    ii. Chief Information Officer
    iii. Chief Operating Officer
    iv. Chief Medical Officer
    v. Chief Financial Officer
    vi. Chief Digital Officer
    vii. Chief Innovation Officer
    viii. Director of Radiology
    ix. Radiologists
    x. Chair of AI
    xi. Chairman
    xii. Dean
    xiii. Professors
    xiv. Residents

- Channels for promotion:
    1. Affiliative Marketing using channel partners like GE health, Philips etc. and conferences like RSNA, JPR & SIIM
    2. Linked in
    3. Twitter
    4. You tube
    5. Webinars

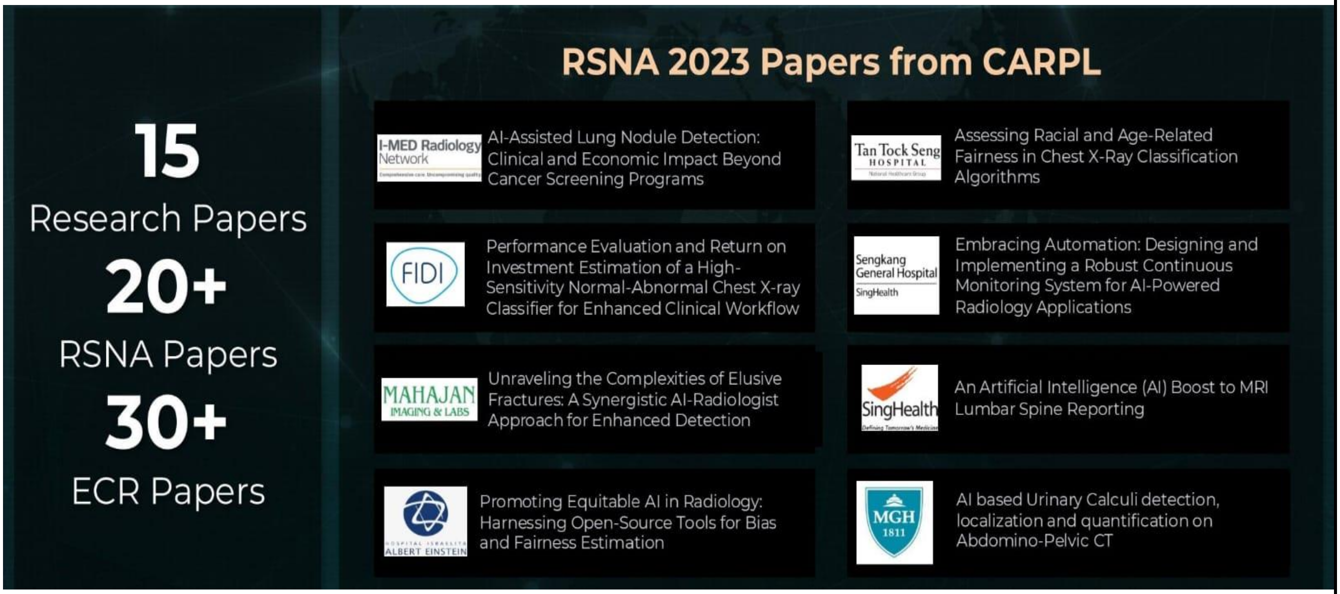
- Objective: To create brand awareness to every HCP and radiologists as well as for the AI enthusiasts. Establishing CTA (Call to Action) the objective of the product should be relevant for the preference of the customer. More focus on persona-based marketing by first identifying the ICP (Ideal Customer Profile) by finding gap and pitch the product by highlighting the benefits.
- Demand should be organically generated:
    1. New Leads
    2. Buzz (Creating FOMO)
    3. Events Conferences
    4. Adding people in Database for continues network creation by cold emails and newsletters for sensitization.

- **Sales Department:**
    - Headed by the CEO
    - Sales is being operated on multiple stages and strategies are blended to communicate and setup meetings to discuss proposals to the HCP based in US AUs
    - The sales process starts from the market research to find the ICP matched professionals and the organizations by generating leads via multiple strategies.
    - **Leads** are being generated by the channels like linked in, organizations websites by searching their leadership team, via articles, research papers, conferences, webinars and B2B finder extensions like Apollo.io.
    - Leads are generated and transferred to database systematically which is operated on Google Sheets.

- Database comprises of the Accounts, Stakeholder Name, Designation, Linked in URL, Linked In status, Email, Comments, AI (if any already deployed).
- Daily 100-200 Leads are fed into the sheet mainly 50% accounts of US and 25% for Brazil & Australia both.
- After this a systematic process of sending linked in requests to the stakeholders with a note from the accounts of CEO, COO, CMO & CTO.
- The status of the requests is also being tracked and continuously updated.
- If the request is accepted for building the network cold emails are transferred to the stakeholder.
- Next stage is **Prospecting** in which after getting response on Linked in or Email from the stakeholder the focus is to setup a formal meeting to discuss the proceedings and introducing CARPL to the client and understanding their prospectives.
- **Qualifying** is the next stage in which information from the client is extracted to understand their requirements and suggesting the right AI solution to fulfil their use case. Along with the Infra requirements & Calculation of Return on Investment (ROI) if they deploy CARPL with their clinical workflow. ROI is generally calculated based on COF i.e. Clinical, Operational and Financial.
- **Unqualified** is the stage where the Organization is not matched with requirements of the ICP proceeding with these accounts generally results in nullifying outcomes.
- The stage in which conversations between CARPL sales team and the clients are ongoing is related to closing the deal and paper-works are generally being initiated like-NDA & SLA this is **Active** stage.
- **Inactive** is the stage where the stakeholders of the organizations start ghosting the CARPL executives.
- **Negotiation** is the stage where almost the deal is locked and discussion related to Infra, timeline and costings are discussed.
- **POC** (Proof of Concept) free trail of the platform along with the AI solution integrated into the client’s clinical workflow
- Won is referred to the accounts which are successfully handed to the Deployment team.
- Lost due to any reason it’s not possible to lock a deal with a particular account.

- All these Stages are managed over a CRM (Customer Relation Management) Software like Fresh sales.

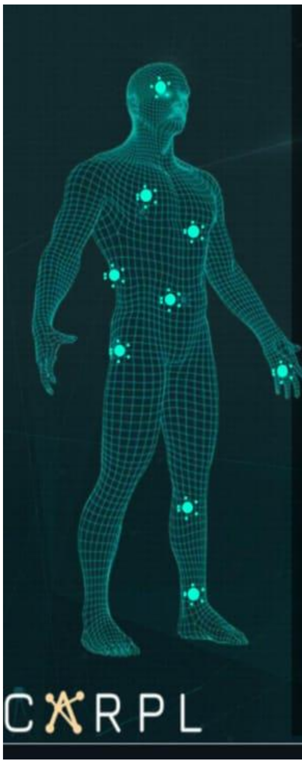
- **Clinical Department:**
  - Headed by the Chief Medical Officer.
  - Indulge in the scientific findings and research works related to the AI, clinical solutions, innovations, population health for better roadmap construction for modification and sustainability.
  - Clinical department have made numerous research projects, and these are being published in research sites and presented in webinars & conferences like RSNA, ECR, SIIM, JPR etc.



- The clinical team every Tuesdays conduct teaching sessions for all the employees on different AI solutions and its workings for better understanding of the working of solutions which can be utilized in their own departments.
- The clinical team also conducts AI's working tests and validation for analyzing the efficiency of the AI and finding the gaps with possible solutions to rectify any issues in the AI functioning. For this they use the datasets from organizations and formulate Confusion matrix along with some descriptive information like sensitivity, specificity, precision and accuracy which further leads to ROC and AUC graphs.

➢ AI- onboarding Department:

- The role of this department is to onboard different AI vendors who all are building AI algorithms for the radiodiagnosis.
- CARPL has collaborated with more than 100+ AI solutions which are over their platform.
- Different Modalities are there like- Xray, USG, CT, MRI &Mamo
- These are used for different body parts and different use cases.
- Legalities and paper works - PIPA



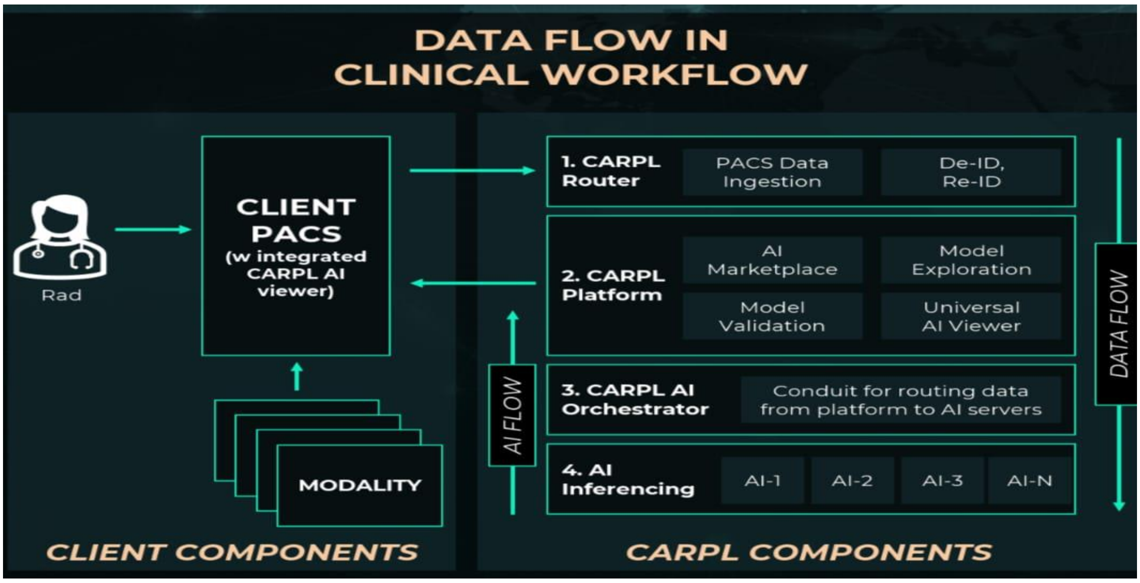
Body Part	Modality	Use-Case	AI companies
Chest	X-Ray	Normal auto-detection	Oxipit, Qure.ai, Lunit, Gleamer, Vuno, VinDr
		Critical finding triaging (PTX, Effusion, Chest tube)	
		Quality Audit	
	CT	Lung cancer screening support	Coreline, Aidence, Vuno, VinDr, Thirona, Aikenist
		Critical finding triaging (PE, Aortic Dissection)	
		Coronary artery calcification	
		COPD quantification	
		Improve quality of low-dose CT	
	USG	Thyroid nodule characterisation	Koios
Breast	2D Mammo	Normal auto-detection	Curemetrix, Medcognetics, Lunit, VinDr
		Cancer detection / second reader support	
		Breast density estimation	
	DBT	Normal auto-detection	Medcognetics
		Cancer detection / second reader support	
	USG	Lesion characterisation	Koios
KUB	CT	Automated renal stone detection, measurement & characterization	BioCliq
Liver	MR	Liver fat and iron quantification	Resonance Health
Prostate	MR	Prostate segmentation, measurement and tumour localisation	Quibim, Ludica Medical
Brain	MR	Brain volumetry for dementia and epilepsy	Icometrix, Quibim, Vuno, Mediaire, In-Med, Aikenist, Subtle Medical
		White matter lesion quantification for MS	
		Automated tumour / aneurysm volumetry	
		Improve quality of fast / low field MRI	
	CT	Critical finding triaging (stroke, infarct, TBI, LVO, fracture)	Avicenna, Qure.ai, IcoMetrix, VinDr
		Automated quantification (bleed volume, ASPECTS score)	
Spine	MR	Automated labeling, reporting and stenosis quantification	Columbo, Synapsica
MSK	X-Ray	Automated fracture detection	Radiobotics, Image Biopsy Labs, AZMed, Vuno, VinDr
		Automated measurements, including OA	
		Automated bone age measurement	

➢ Deployment Department:

- The basic function of this department is to deliver what has been promised by the sales department to the client and expansion.
- Expansion is pitching and offering AI algos and suggestions to the client for using different solutions which can result in more benefit in their clinical workflow.
- Deployment department comprises of Project managers and the technical experts.
- The project managers design the project workflow with respect to timeline and the client requirements by various meetings and constructing roadmaps.
- They must build a robust path right from understanding the Infrastructural designing

at the client’s level ensuring smooth integration of platform as well as seamless AI inferencing over the scans.

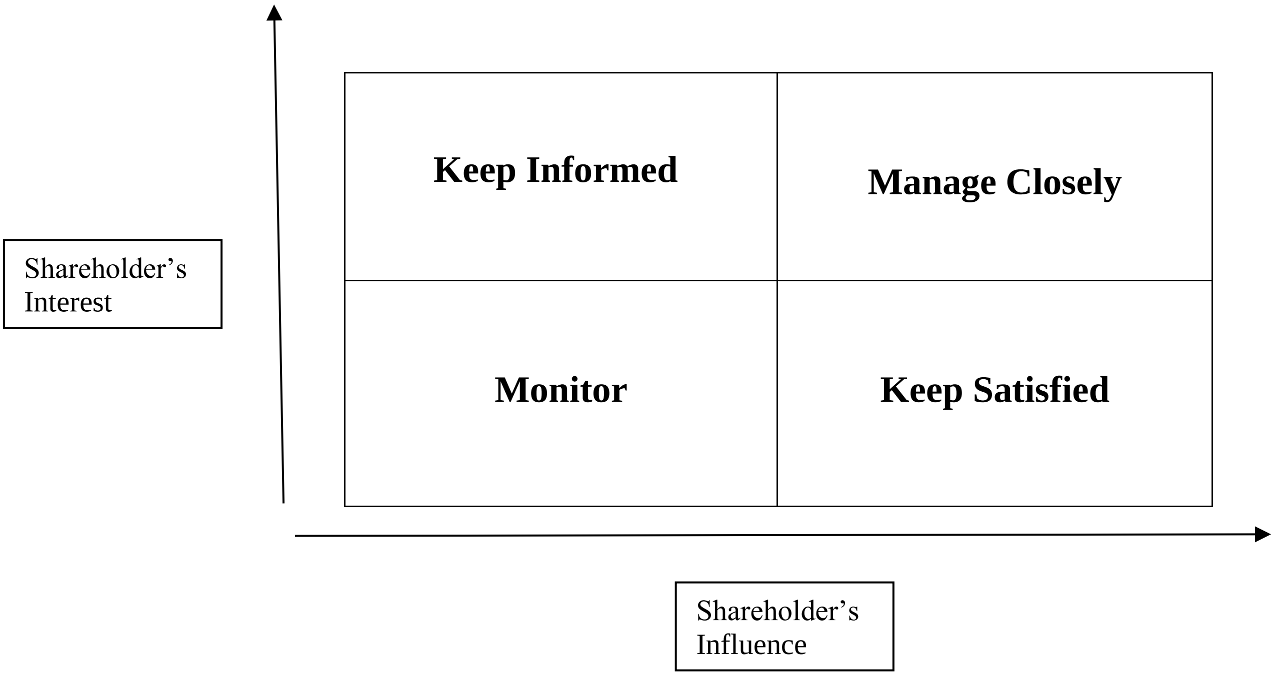
- They need to discuss and understand the requirements of the Client and strategies are made accordingly.
- AI can be deployed through these provisions:
  i. On premise: On-premises deployment refers to the installation and running of software applications on the hardware and infrastructure that are located within the physical premises of an organization, such as in their own data centers or server rooms.
  ii. On cloud: Cloud deployment refers to the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the internet ("the cloud") to offer faster innovation, flexible resources, and economies of scale.
  iii. AI vendor Infra: All the services are provided over the infrastructure of the AI vendor and CARPL platform just works as the bridge for transmitting the scans from the Clients PACS to the infra of AI vendor.
- The Router needs to be on the Clients infra irrespective of the agreed mode of deployment because the priority of anonymizing the data before the leaving the clients infra to avoid any breach
- CARPL platform can be deployed On- prem or on cloud
- CARPL orchestrator is for routing the scans from platform to the AI servers for inferencing also overlooks on the essential DIACOM tags required for the AI to run over the scans.
- CARPL orchestrator can be also deployed on both on prem or on cloud.
- AI inferencing can be deployed on cloud, on platform or on the AI vendor information



SECTION 4: LEARNINGS -

Sales department

- ✓ The complete procedure of finding different stakeholders of HCPs
- ✓ Strategies like finding HCPs where our competitors are deployed like Blackford, AI DOC, PIN AI, Rad AI and Ferrum.
- ✓ Recent articles on AI by radiologists or professors to build network.
- ✓ management of email and LinkedIn requests campaign along with tracking
- ✓ Updating the status of accounts using CRM software
- ✓ Stakeholder mapping:
  - It is a type of metrics which shows the importance and roles of the different stakeholders associated to the business.
  - Helps to appropriately manage and inform your stakeholders.
  - Helps in identifying and organizing key stakeholders.
  - The axis of the metrics: Influence and Interest.
  - Influence of the stakeholder over the business projects through funding and legal compliance etc.
  - Interest is how much the stakeholder is engaged with the project either help the process or hurt it.

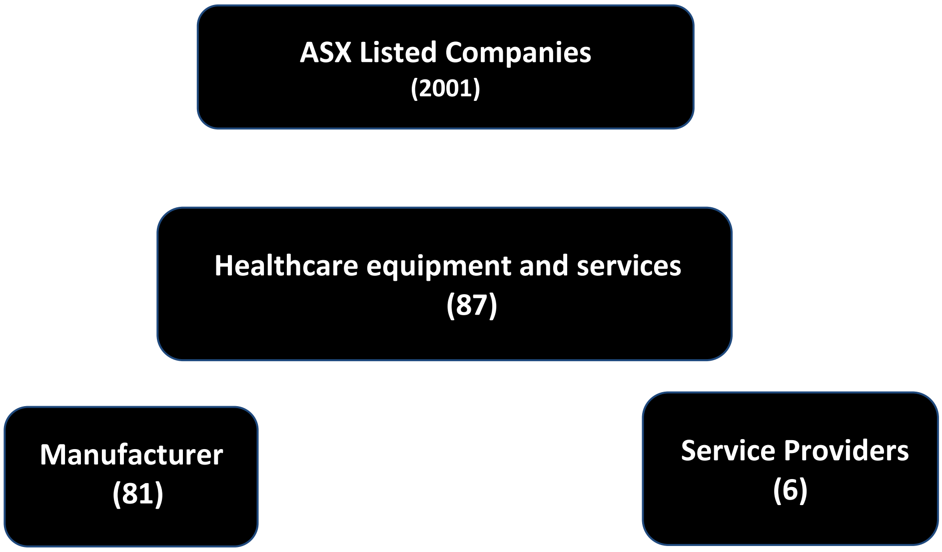


- Manage Closely: High Influence and high interest. They could make and break the project so it’s important to give regular updates and ensure if they need something or not.
- Monitor: Low Influence and low interest. Should send regular updates and monitor for any change in influence and the interest but not to worry too much about them.
- Keep Satisfied: High Influence but low interest. Priority should be on identifying their preferences as they do not have much interest in the projects so regular updates won’t work for them one needs to handle delicately.
- Keep informed: Low influence but high interest. These are the stakeholders who care about the outcome of the project but do not have control over its success so just keep them updated with the projects.

✓ Use of extensions to extract relevant information from websites and LinkedIn like Appolo.io

✓ Analytical approach application - example task of finding the profit gained by the different radiological public companies of Australia per spine MRI.

1. First, we found the ASX listed companies and breakdown the radiology service companies



So, the six Radiological Service providers listed are:

- Integral
- Lumus
- Capitol
- Quantum
- Sonic health
- Imagion

2. We referred to the annual reports of these companies and compile the annual profit generated by the Radiological diagnosis.
3. Then we assumed by the market trends the number of MRI scans are 15% of the total scans and 35% of MRI are Spine MRI
4. 40% of the total profits from scans are of MRI scans and further 40% of the MRI scans are from the spine MRI.
5. Divide the annual scans of spine MRI by 260 (that's the number of working days in a year)
6. Now we calculated the per day profit of the spine MRI by dividing the annual from 260.
7. At blast multiply the Per day profit from the number of daily spine MRI scans to get

the Profit per spine MRI scan.

8. We did this for three companies:
    - Integral: $38.89
    - Helius: $19.40
    - Capitol: $40.50

- ✓ Effective use of LinkedIn sales navigator to find prominent match to ICP
- ✓ AI innovations and solutions which are being integrated in the clinical departments which hold the ability revolutionize the health system.

Marketing

- ✓ Concepts of Platform as a service and Software as a service
    - PAS: Is a cloud computing service model that provides a platform for developers to build, deploy, and manage applications. It eliminates the need to manage the underlying hardware and software layers, allowing developers to focus on writing code and developing applications.
    - SAS: Is a cloud computing service model that delivers software applications over the internet on a subscription basis. Users can access and use these applications via a web browser, without needing to install or maintain the software locally.

- ✓ How to use different channels to create brand awareness and tracking the performance like YouTube views, subscribers, impressions on cold emails, clicks on newsletters, webinar registrations and attendees, LinkedIn followers as well as post impressions
- ✓ Handling and continuous improvement of website to give better user experience
- ✓ Importance of feedback from the clients and use the positives to market and communicate the gaps to the authorities
- ✓ From conferences and webinars scraping information of attendees and build connection
- ✓ Strategies like to increase the brand value encourage the employees to develop connections and stay active on LinkedIn
- ✓ Strategic partnerships is another channel for marketing in which companies who are into developing PACS and RIS manufacturers like -GE & Philips. They deliver the PACS to the client along the installed CARPL platform. Also share booth space in conferences and seminars which increase CARPL's visibility and awareness.

Clinical Department

✓ Weekly AI classes helped in gaining so much about different AI solutions:

1. MSK AI solution for Bone Age calculation:

So, there are two types of Age- One is chronological age which is according to the DOB and other one is the bone age. Ideally bone age is needed to be parallel to the chronological age but due to any abnormality bone age delays or exceeds. Increment in bone age at an early age cause early fusion and conversion of cartilage into bones & the growth points (Ossification centers) stop to grow further that’s why the children who are unusually taller in early years eventually remains shorter in the long term.

The mid-parental height can be calculated by the formulae:

$$\frac{\text{Father's Height} + \text{Mother's Height}}{2} + 6.5 \pm 10$$

AI companies providing these solutions are: Vuno and Panda

So, there are 2 methods using which AI calculate the bone age:

- GP (Greulich Pyle): Asses the ossify centers of bones in the hand and wrist to estimate the bone age and uses the Atlas for referencing.
- TW (Tanner Whitehouse) version 3: Also known as RUS (Radius, Ulna and short bones). Evaluation of each bone and according to bone’s maturity total bone age estimation.
- TW 3 is used in Sports Authority of India.

2. Cina Chest CT

Specifically used for the Triaging of patients. Generally, the solution runs on contrast CT in which some contrast agent is injected for increasing the visualization clarity of the CT scans. It helps in finding pulmonary embolism cases and aortic dissection. In pulmonary embolism there is a plug or embolism may be of fat or air clot in the arteries supplying blood to the lungs and Aortic dissection is the case in which due to hypertension and aneurysm of heart the aorta’s inner wall ruptures and the blood clots in the region between outer and inner wall of the aorta.

3. CT Head:

3 solutions used for the CT head CINA ICH, CINA Aspects and CINA LVO. All these are used for Triaging of the cases. The ICH and Aspects work on NCCT head in which ischemia

and hemorrhage are being detected respectively. They measure the severity based on the Hounsfield unit which is used to measure the Volex values in CT. If the HF is nearer to the +1000 may be the case of hemorrhage and if -1000 ischemia is the possible reason.

-Testing of AI efficiency:

In this the procedure of testing the AI's performance was learnt. The basic tools which are used for the AI testing – Confusion Matrix, Re-operative Curve and AUC curve.

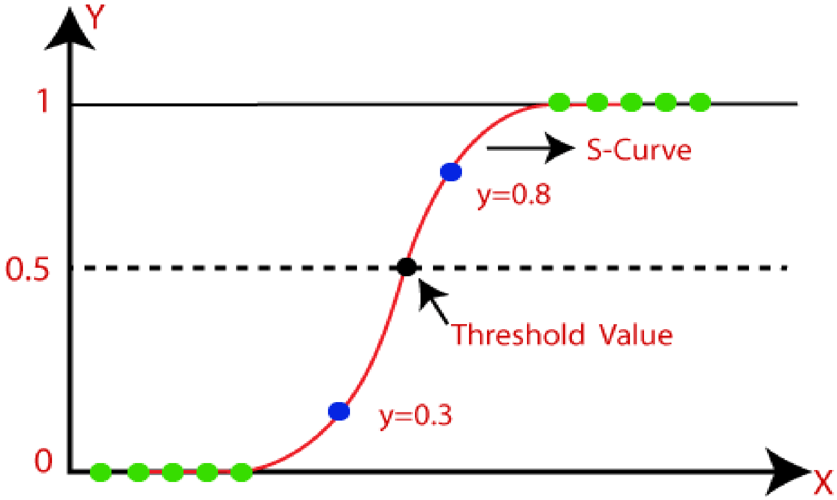
- So, suppose the AI functions for classification Disease & No Disease.
- AI provides the Output in the form of 0 to 1, by imposing the sigmoid function for logistic regression for getting predictive probability of the presence of disease and no disease. The formula of sigmoid function is:

$$S(x) = \frac{1}{1+e^{-x}}$$

S(X)= probability of dependent variable based on independent variable

X=Independent Variable

- Probability nearer to 1 predicts higher possibility of disease and nearer to 0 is non diseased condition.
- These probabilities are plotted in Logistic regression, and we decide the threshold level which converts the probabilities to a complete Boolean value suppose 0.5 is the threshold then probabilities above the 0.5 will be considered as 1 and below 0.5 as 0.



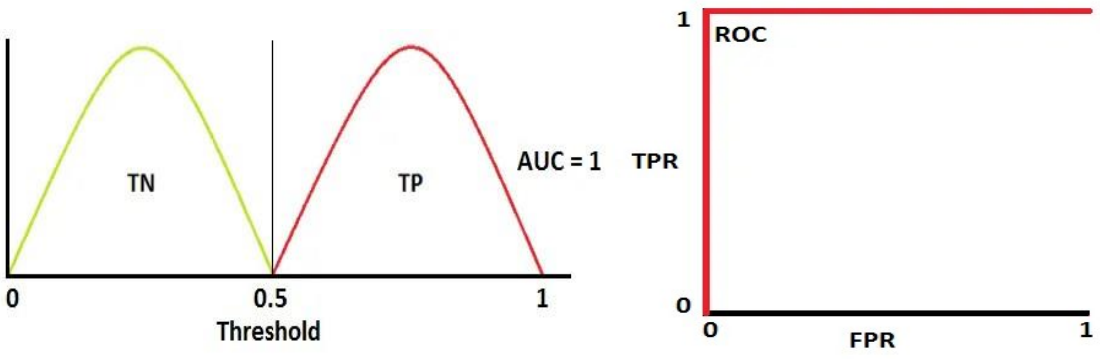
Logistic Regression plot

- Now we use Ground Truth which is the conversion of the interpretations from the Radiologists into Boolean values for comparison with the AI outputs
- Using GT and the AI outputs construct the Confusion Matrix.

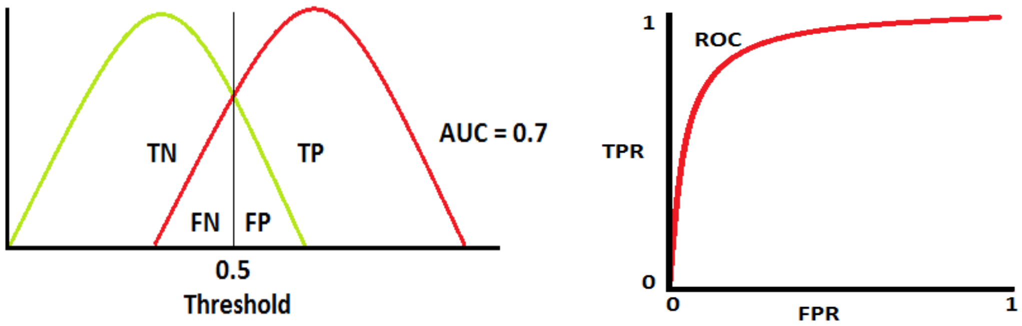
Predicted Class

		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

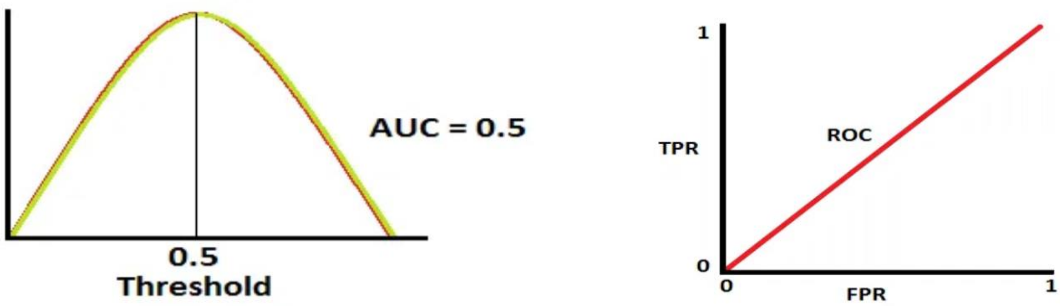
- TP: When the patient is actually diseased, and AI output the claims the same.
- TN: When the patient is not diseased, AI too claims the same.
- FP: When AI claims that's the patient is diseased, but patient is not diseased.
- FN: When AI claims patient does not have any disease, but the patient is diseased.
- ROC (Receiver Operating Curve): By calculating the TP Rate and the FP Rate this can be plot for finding the ability of the AI model to classify diseased and non-diseased.
- AUC: It's the area covered under the ROC



So, this is the ideal situation in which the two curves of probabilities do not overlap at all, and the AUC is 100% which means that the model has the 100% ability to distinguish between diseased and non-diseased.



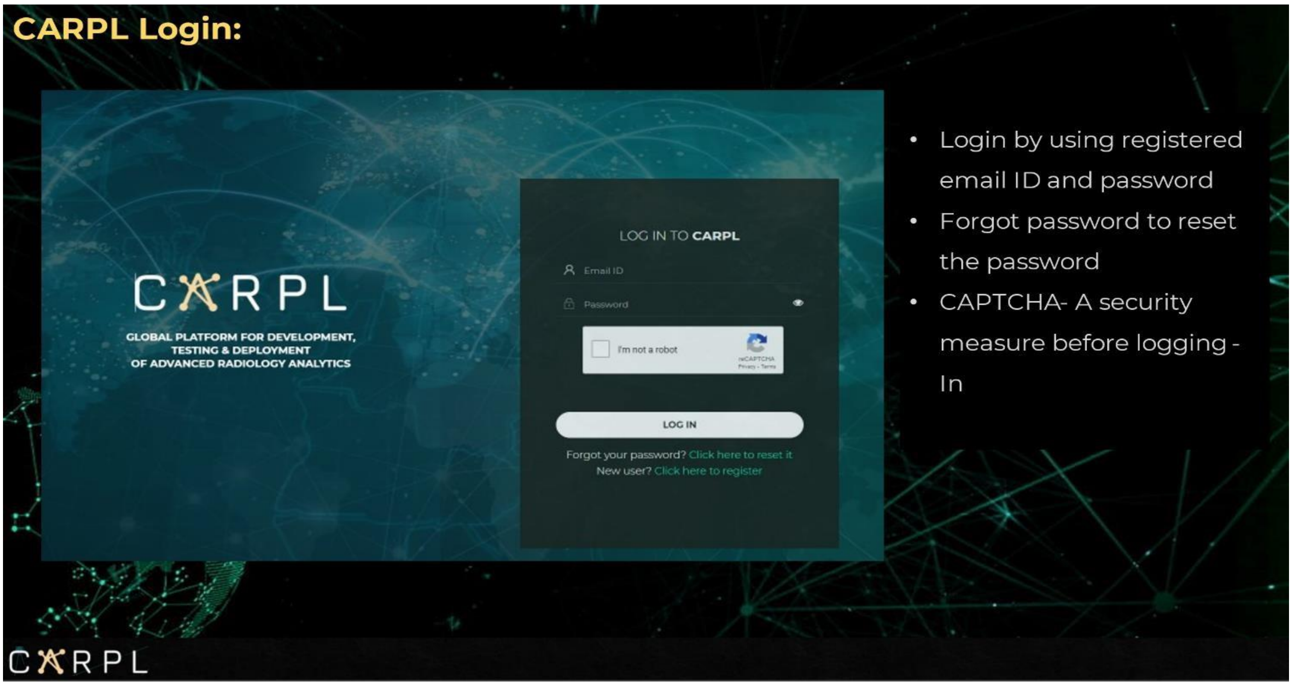
When two distributions overlap, there are type 1 and type 2 errors. When AUC is 0.7, it means there is a 70% chance that the AI model will be able to distinguish between diseased and non-diseased.

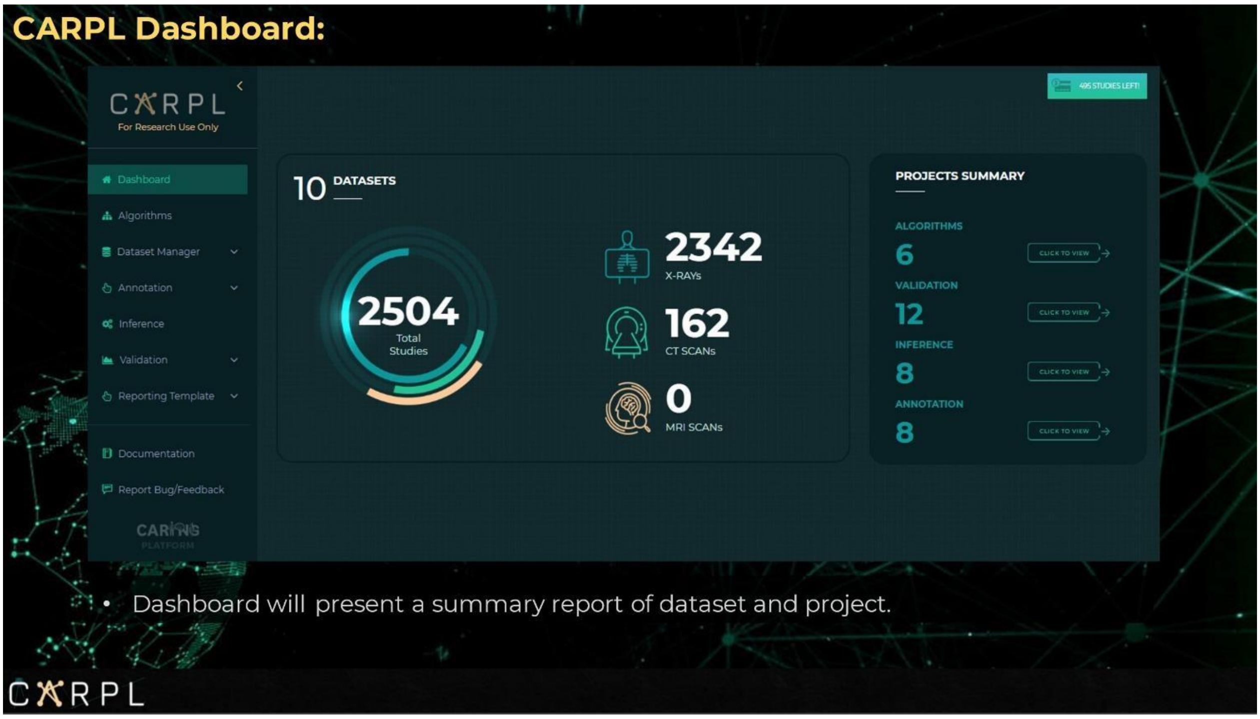


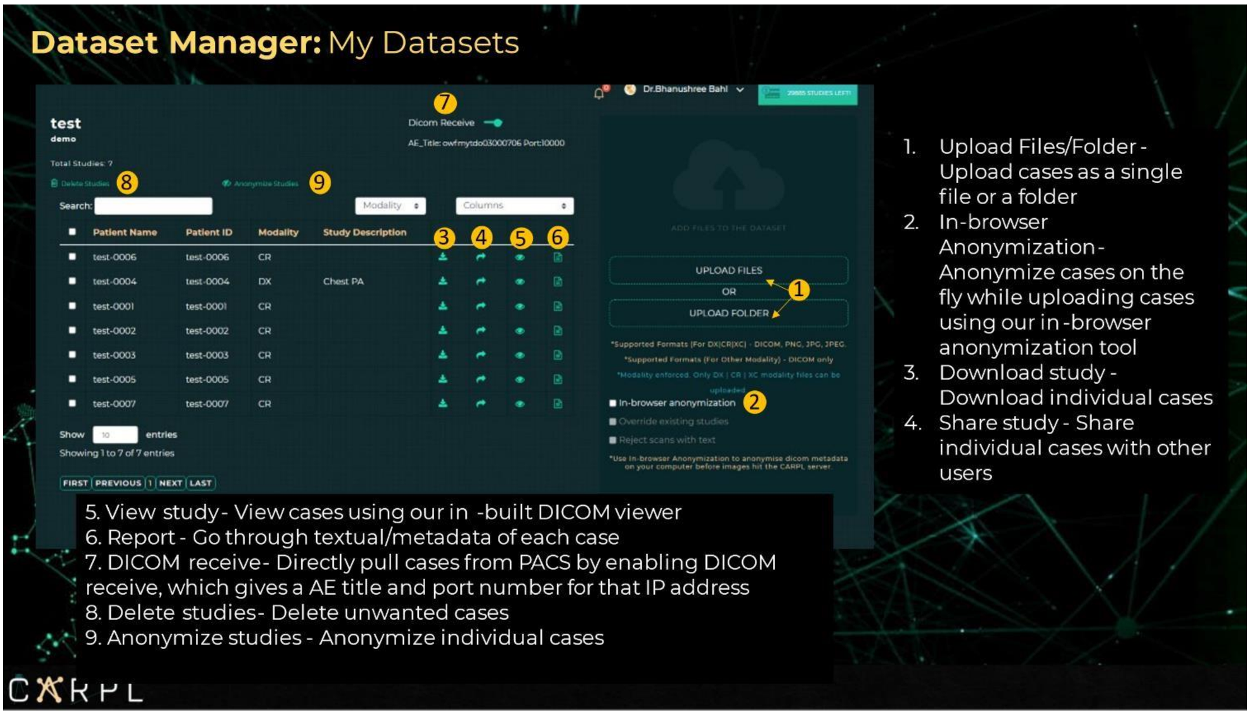
This is the worst situation. When AUC is approximately 0.5, the model has no ability to distinguish between positive class and negative class.

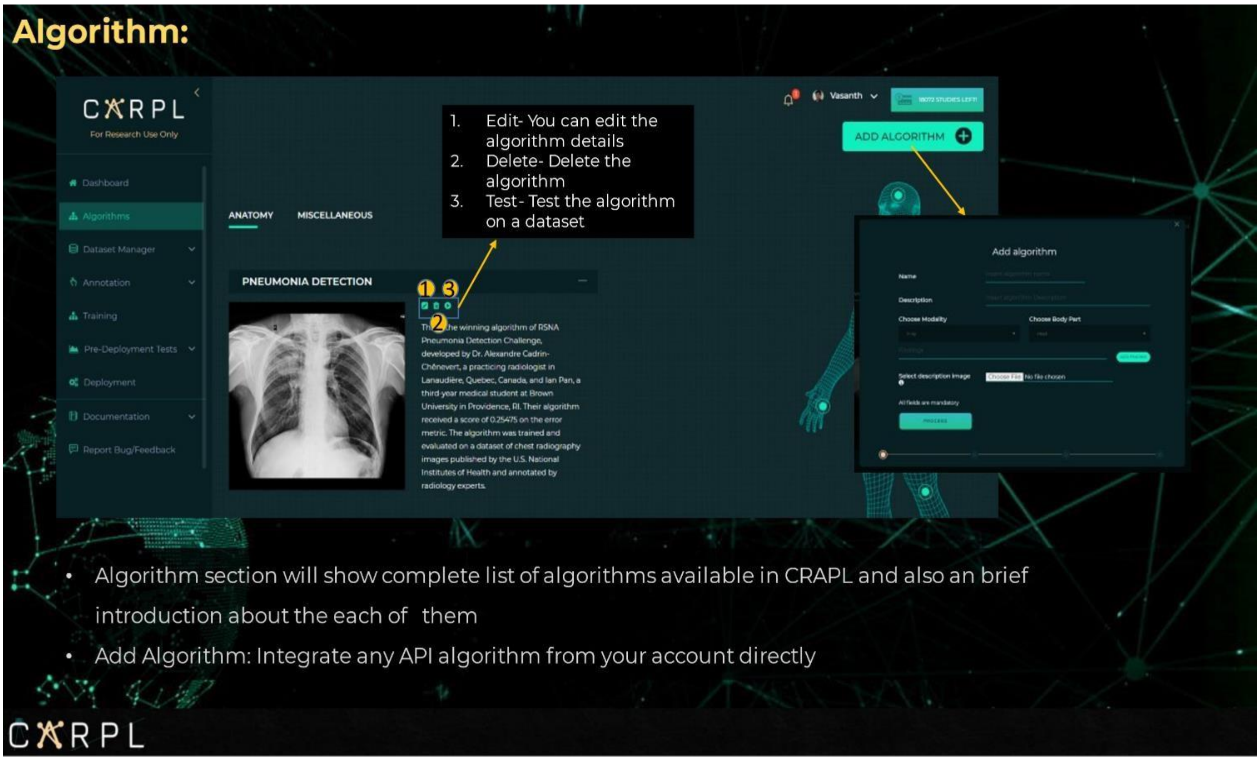
Tech Department

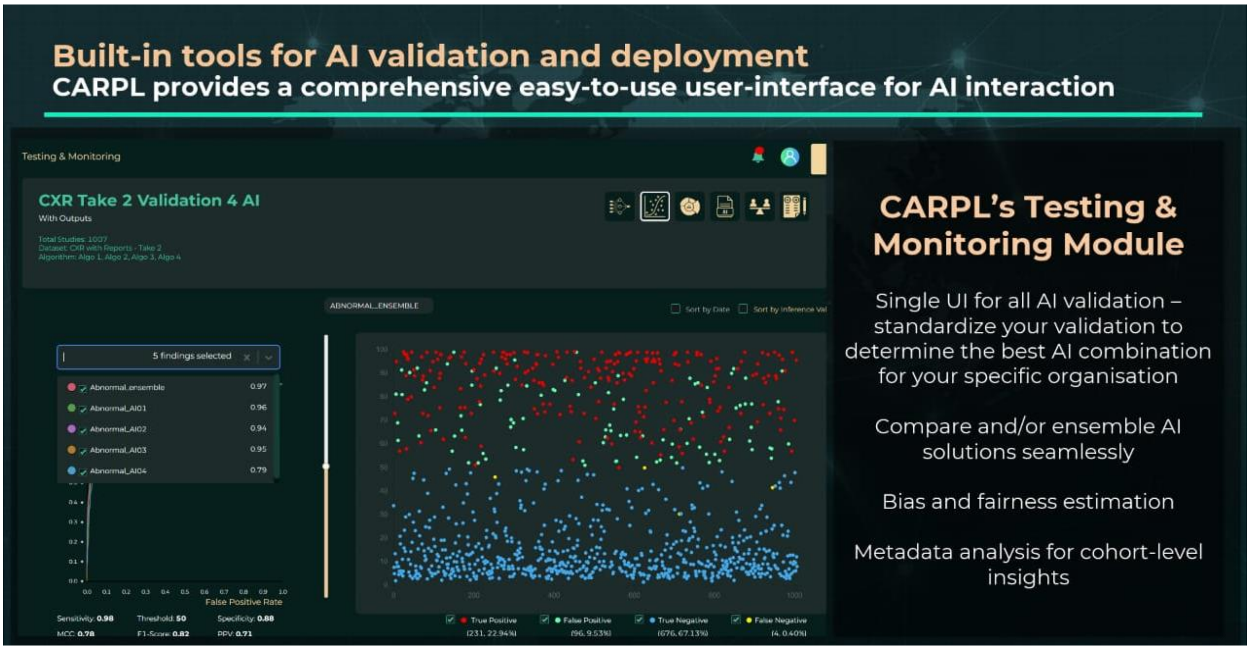
I got the experience of Platform interface and its usage:





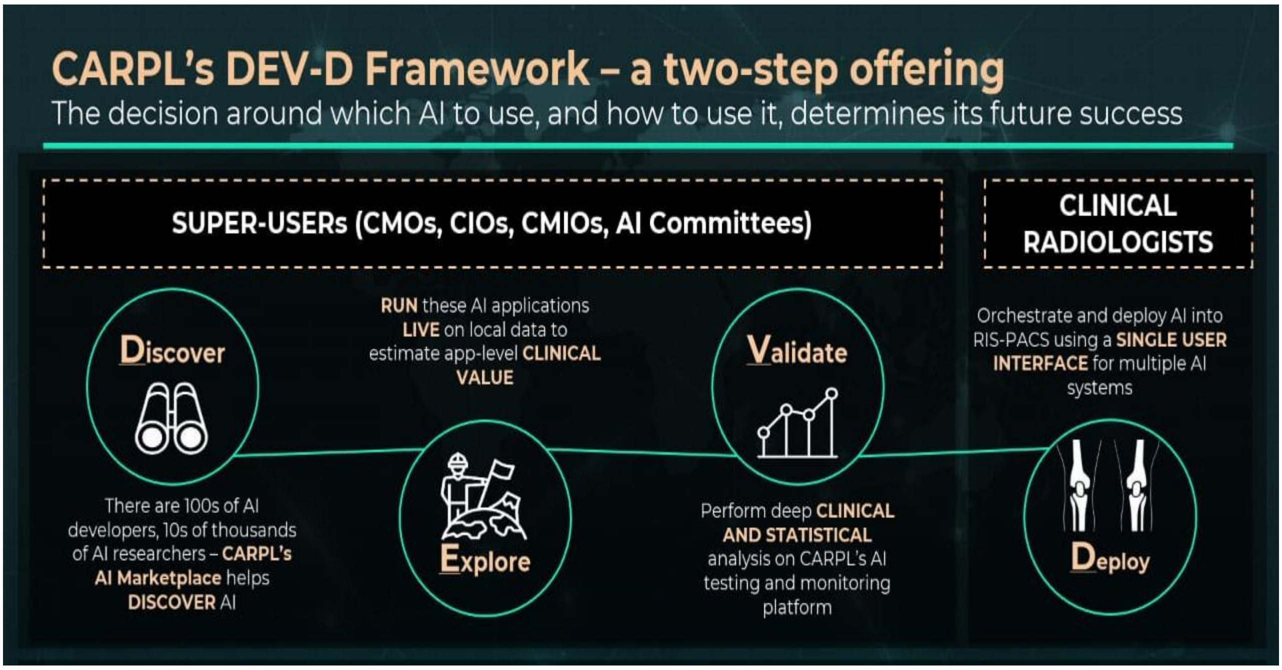

Dataset Manager: My Datasets
1. Upload Files/Folder - Upload cases as a single file or a folder
2. In-browser Anonymization - Anonymize cases on the fly while uploading cases using our in-browser anonymization tool
3. Download study - Download individual cases
4. Share study - Share individual cases with other users
5. View study - View cases using our in-built DICOM viewer
6. Report - Go through textual/metadata of each case
7. DICOM receive - Directly pull cases from PACS by enabling DICOM receive, which gives a AE title and port number for that IP address
8. Delete studies - Delete unwanted cases
9. Anonymize studies - Anonymize individual cases
UPLOAD FILES
OR
UPLOAD FOLDER
CARPL


Algorithm:
1. Edit - You can edit the algorithm details
2. Delete - Delete the algorithm
3. Test - Test the algorithm on a dataset
ADD ALGORITHM
Add algorithm
PNEUMONIA DETECTION
ANATOMY
MISCELLANEOUS
• Algorithm section will show complete list of algorithms available in CRAPL and also an brief introduction about the each of them
• Add Algorithm: Integrate any API algorithm from your account directly
CARPL



Along with the platform the concept of DEV-D acts as the differentiating factor of CARPL. The DEV -D model stands for **Discover, Explore, Validate and Deploy.**

- **Discover:** More than 100+ AI solutions which are being onboarded on the platform which are based on different use cases like part under the study, modality used for the part's visualization and findings in the scan. For discovering client can click on the algorithm section to know everything about the AI algorithm.
- **Explore:** Now after discovering the AI solutions according to the HCP, they can choose the solution to test the model by a small data set of their own healthcare setup. On platform one can visit the deployment section, upload their desired dataset and run the solution over the platform.
- **Validate:** Now this the USP of CARPL, it provides the provision to the user for validating the AI results on the large dataset. Users can visit Testing & Monitoring and upload the data and this module will produce Confusion matrix, AUC and ROC for the analysis of the AI's functioning.
- **Deployment:** This is the final when all the agreements the platform is integrated with the PACS – RIS of the client.



Deployment Department:

The complete procedure of one deployment project in India for TB screening program in rural and remote areas with the collaboration of CHAI (Clinton Heath Access Initiative) along with Lab India. Lab India provides the Hardware setup -Portable Xray device, Tripod and Laptops etc. CARPL provides the RIS-PACS integrated platform with Qure.AI Chest Xray solution.

In the program these procedures are needed to be done:

1. Camp registration
2. Beneficiary Registration
3. TB screening
4. Sputum collection
5. Post camp sputum test/ Follow ups

As these setups are required to be executed in remote area connectivity and internet connections are the major concern due to which two servers are created:

1. Central PACS & RIS
2. Local PACS & RIS

So, the officials can store the information at the local server and all the data when connected to the internet gets synced to the central cloud-based server.

RECOMMENDATIONS:

- Before onboarding the AI vendor’s AI solutions on the platform CARPL should conduct AI testing and evaluation for providing quality services to the clients.
- Collaborate with more hospitals and healthcare institutions to access diverse and larger datasets, enhancing the training and validation of AI models for generalizability.
- Should expand to other horizons like UAE, UK& Japan etc.
- Develop comprehensive training programs for healthcare professionals to effectively use AI tools and understand their outputs.
- Work on seamless integration of AI solutions with existing healthcare IT systems like Electronic Health Records (EHR) to streamline workflows.
- Invest in R&D to explore new applications of AI in healthcare beyond radiology, such as pathology, cardiology, and personalized medicine.
- Develop resources and tools to educate patients about how AI is being used in their care, addressing any concerns about accuracy and privac

PROJECT REPORT

SECTION 1: INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative force in healthcare, promising to revolutionize medical practice, patient care, and healthcare management. By providing advanced algorithms and computational power, AI technologies are competent enough to enhance diagnostics, personalize treatment plans, streamline administrative tasks, and improve overall healthcare outcomes. From analyzing vast amounts of medical data to supporting clinical decision-making, AI is reshaping how healthcare providers deliver services and how patients experience care. In recent years, AI applications have demonstrated remarkable potential across various healthcare domains, ranging from radiology and pathology to genomics and personalized medicine. Machine learning algorithms can go through immense datasets to identify patterns, predict outcomes, and assist in early disease detection. This capability not only speeds up diagnosis but also enhances accuracy, thereby reducing errors and improving patient safety. This study is basically based on the integration of Radio-diagnosis with AI so, the basic workflow of any healthcare institution with radiology services comprises of:

- **Registration**

  Beyond clinical applications, AI is also revolutionizing healthcare operations and administration. Administrative tasks such as scheduling, billing, and patient record management can be automated with AI systems, allowing healthcare providers to focus more on patient care and less on paperwork. Furthermore, predictive analytics powered by AI can forecast patient demand, optimize resource allocation, and improve operational efficiency within healthcare organizations example Xsolis AI which integrates with EHR of the institution.

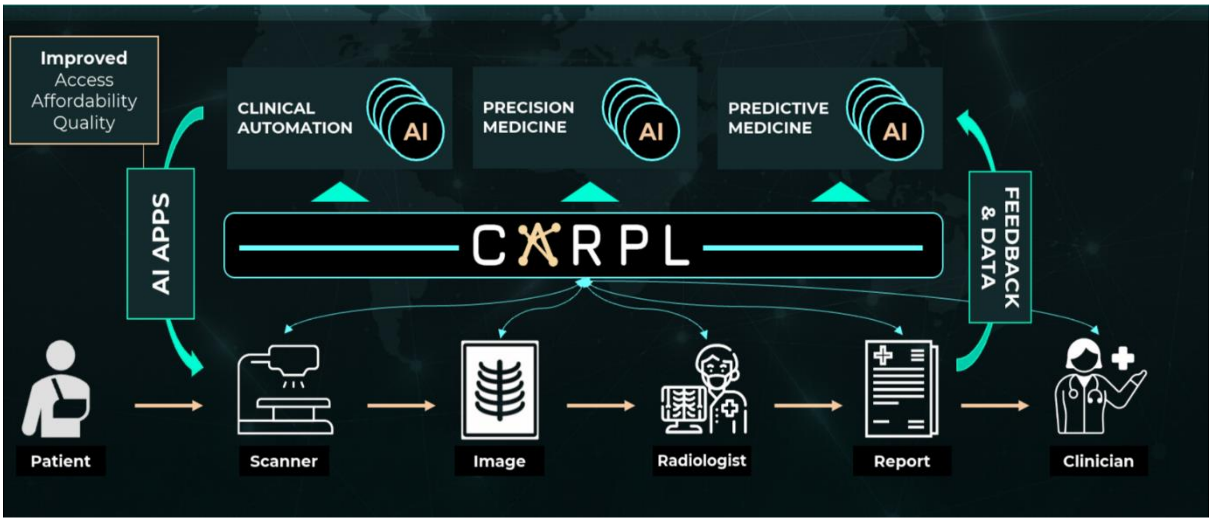
- **Image Processing**

  AI models are incorporated into the diagnostic machines like X ray, Ultrasound, CT, MRI and Mamo as well as in the viewers application for image visualizations guiding the technicians for better image acquisition by effective contrast management analysis and to enhance clarity of the image being acquired example Subtle imaging AI company.

- **Reporting**

  AI models are also incorporated with the PACS and RIS which helps in Scan interpretations, triaging and assistance to the Radiologists for report formation and better clinical diagnosis and correct intervention. Example CARPL incorporated into the PACS RIS of HCPs and run AI models on different modalities for different use cases.

AI ENABLE CLINICAL WORKFLOW



In the field of radiology, AI algorithms are increasingly being integrated to assist in the interpretation of medical imaging, particularly in musculoskeletal (MSK) X-rays. This integration aims to enhance diagnostic accuracy, reduce human error, and optimize workflow efficiency.

Our research focuses on the application of AI in analyzing MSK X-rays. We have compiled a dataset consisting of 4336 cases from a US based hospital, encompassing various body parts such as the leg, ankle, and foot. For each case, we have meticulously calculated the ground truth (GT) and compared it with the AI-generated findings. This comparison has enabled the development of confusion metrics, and the calculation of key performance metrics including specificity, sensitivity, and precision.

In addition to the performance analysis, we have extended our study to assess the practical impact of AI-assisted diagnosis. We obtained a secondary dataset from the same hospital, containing radiological reports interpreted with and without AI assistance. This dataset also includes interpretations by junior and senior radiologists. By analyzing the concordance and discordance between these reports, we aim to determine the statistical significance of the differences in diagnostic outcomes. With AI interpretations are 329 cases of lower extremities X-ray reports and without AI 334 cases.

Through this comprehensive analysis, our research seeks to elucidate the efficacy of AI in MSK radiography and its potential to augment the diagnostic capabilities of radiologists. Ultimately, this study aims to provide valuable insights into how AI can be effectively integrated into clinical practice to improve patient care.

1.1 RATIONAL OF STUDY

Rationale:

The integration of Artificial Intelligence (AI) into radiology has shown potential in improving diagnostic accuracy and efficiency. This study evaluates the practical application of AI in detecting fractures in leg, ankle, and foot radiographs and its impact on diagnostic concordance between junior radiologists (JR) and senior radiologists (SR). Insights from several research articles underscore the potential benefits and limitations of AI in this context.

Improvement in Diagnostic Accuracy:

AI systems have demonstrated high accuracy in image recognition tasks, including fracture detection. For instance, a study highlighted that AI increased the detection accuracy of fractures by 15-20%, which can support radiologists in making more precise diagnoses. Specifically, Lindsey et al. (2021) found that AI assistance led to a 22% improvement in fracture detection sensitivity.

Reduction of Discrepancy Rates:

Discrepancy rates between preliminary and final radiology reports are a significant issue, often leading to delays in treatment and adverse patient outcomes. AI can play a critical role in minimizing these discrepancies by providing consistent and reliable preliminary interpretations. For example, a study reported that AI reduced the discrepancy rate by 25%.

Brady et al. (2012) emphasized that performance indicators and quality improvement methods, which include AI, could significantly lower discrepancy rates by up to 30% .

Efficiency in Clinical Workflow:

AI can streamline clinical workflows by automating routine diagnostic tasks, allowing radiologists to focus on more complex cases. This efficiency gain not only speeds up the diagnostic process but also enhances the productivity of radiology departments. AI integration has been shown to reduce the time taken for diagnostic reads by 20-30%, thereby increasing the throughput of cases processed . Tang et al. (2021) noted that AI could lead to a 25% increase in radiologist productivity.

Addressing Limitations and Bias:

Despite its potential, AI has limitations. One concern is the possibility of AI misinterpreting non-fracture abnormalities as fractures, leading to false positives. This includes conditions like curvilinear densities, effusions, degenerative changes, and other clinical abnormalities. Continuous improvement in AI systems is necessary to mitigate such risks and enhance diagnostic accuracy. Scheinfeld et al. (2021) highlighted that improving AI algorithms could reduce false positive rates by 10-15% .

NEED FOR THE STUDY:

Given the promising results and challenges associated with AI in radiology, this study is essential to:

Evaluate Real-world Application: Assess the practical use of AI in clinical settings, focusing on its impact on diagnostic concordance and discordance between JR and SR radiologists.

Identify Benefits and Risks: Highlight the potential benefits of AI assistance in reducing diagnostic discrepancies while identifying risks associated with its use, such as misclassification of certain clinical conditions.

Provide Evidence for Integration: Generate empirical evidence to support the integration of AI in radiology, guiding future implementations and improvements in AI algorithms to better serve clinical needs.

This study aims to contribute to the growing body of literature by providing empirical data on the effectiveness and limitations of AI in radiology, ultimately aiding in the optimization of

AI-assisted diagnostic tools for improved patient care. By building on insights from previous studies, this research will help in understanding the practical implications of AI in enhancing radiological practices.

1.2 RESEARCH QUESTION

Primary Question:

How does the diagnostic performance of AI algorithms in interpreting musculoskeletal (MSK) X-rays compare to the ground truth for the lower extremities?

Secondary Question:

Is there a statistically significant difference in concordance and discordance between the interpretations of X-ray scans of lower extremities by senior radiologist and junior radiologist with AI-assistance & non-AI-assistance?

1.3 OBJECTIVES

Primary Objective:

To evaluate the diagnostic performance of AI algorithms in interpreting musculoskeletal (MSK) X-rays by comparing AI findings with the ground truth (GT) and calculating key performance metrics such as specificity, sensitivity, and precision for various body parts (Leg, ankle, and foot).

Secondary Objectives

1. To evaluate the concordance and discordance between interpretations of Xray scans of lower extremities by Senior radiologist and Junior radiologist with AI-assistance & non-AI-assistance.
2. To identify statistically significant differences in diagnostic outcomes between AI-assisted and non-AI-assisted interpretations.
3. To provide insights into the practical application of AI in clinical settings and its potential benefits and limitations for patient care.

SECTION 2: MODES OF DATA COLLECTION

Secondary data has been provided for the retrospective cross-sectional analysis of the MSK X-ray reports interpreted by the radiologists of a US based hospital in which a MSK X-ray AI solution is deployed for detecting fractures and triaging. Along with the Radiologist interpretations there were the AI findings too with the reports.

There were 4,363 X-ray reports out of which 255 cases where AI fails to give any outcome and 4108 cases where AI successfully provided output. These reports were in csv format and the AI findings were in the categorical format, **Suspicious Finding**- Presence of Fracture & **No Finding**- Absence of Fracture.

For further analysis and secondary objective study from the same hospital 663 MSK X-ray reports were provided which compiles of the interpretations of both the junior radiologist and the senior radiologist. Out of these 663 the junior radiologist is assisted by the AI in 329 cases whereas in 334 cases no AI assistance was being used.

The data which was provided is highly confidential as contains the patient's information along with the complexities so ethically the name of the hospital and the name of the AI developer company is not used for this report.

SECTION- 3: METHODOLOGY

3.1 <u>DATA COMPILATION</u>

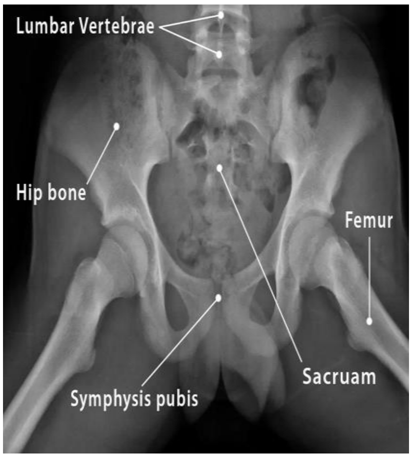
The data was provided in csv format and MS excel was used for compiling the data and csv was converted into columns by removing the delimiter. Structured columns of the data were created that are –

- **Group name**: The group consists of different body parts or comprises different bones present within it. There are Shoulder, Arm, Elbow, Forearm, Wrist, Hand, Leg, Ankle and Foot. Example: Leg comprises of Bones like Hip, Femur, Patella, Tibia and Fibula.
- **MRN number**: It's the unique ID provided to each patient undergoing the X-ray test.
- **Accession number**: It's the unique ID given to each X-ray scan means the same patient having X MRN number if undergone X-ray of shoulder with 2viwes and also X-ray of Foot with 2 views then 4 accession numbers will be allotted to his scans.

➢ **Body Part**: These are the bones which will be present in X-ray scan of the above-mentioned groups. Examples are.

Shoulder

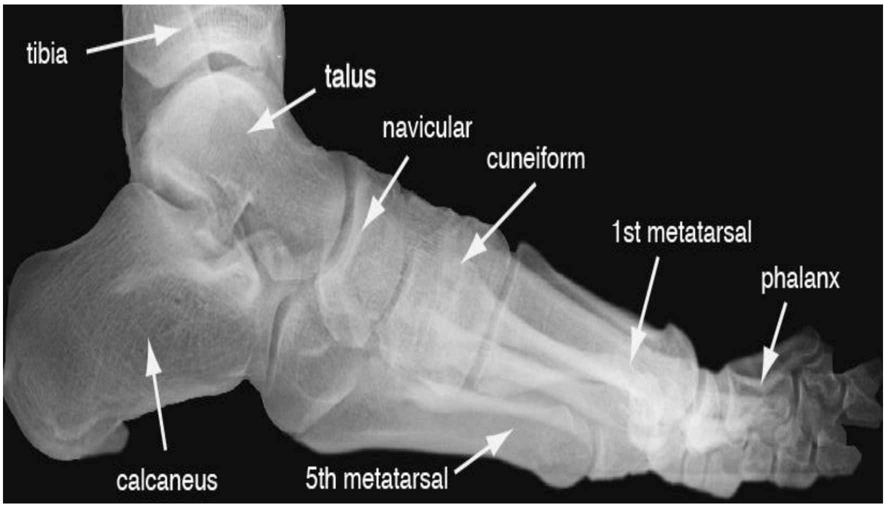
Elbow



Hip

Leg

Ankle



Foot

(These X-ray scans are provided for reference to understand which bone comes under which group)

- **Impression:** It’s the interpretation of the X-ray scan by the Radiologist which will be used to create GT.
- **AI Findings:** It’s the finding of the AI which was converted from the suspicious finding to 1 and no finding to 0.

	A	B	C	D	E	G	H
1	Group	MRN	Accession	BodyPart	Impression	COMMENTS	AI
18	LEG	95660	40340490	BN FEMUR : MIN 2 VIEWS	1. NORMAL RADIOGRAPHS LEFT FEMUR AND LOWER LEG. I PERSONALLY REVIEWED THE IMAGES/STUDY AND I AGREE WITH THE RESIDENT FINDINGS AS STATED. THIS STUDY WAS INTERPRETED AT UNIVERSITY HOSPITALS CLEVELAND MEDICAL CENTER, CLEVELAND, OHIO.		0
19	LEG	95660	40340491	BN TIBIA	1. NORMAL RADIOGRAPHS LEFT FEMUR AND LOWER LEG. I PERSONALLY REVIEWED THE IMAGES/STUDY AND I AGREE WITH THE RESIDENT FINDINGS AS STATED. THIS STUDY WAS INTERPRETED AT UNIVERSITY HOSPITALS CLEVELAND MEDICAL CENTER, CLEVELAND, OHIO.		0
34	LEG	289936	39849541	BN BILATERAL FEMUR : MIN 2	1. NO ACUTE FRACTURE OR MALALIGNMENT OF THE RIGHT OR LEFT FEMUR. 2. POSTOPERATIVE CHANGES OF BILATERAL TOTAL KNEE ARTHROPLASTIES WITHOUT HARDWARE COMPLICATIONS. 3. NUMEROUS RETAINED BALLISTIC FRAGMENTS PROJECT OVER THE PELVIS AND BILATERAL PROXIMAL FEMORAL SOFT TISSUES. I PERSONALLY REVIEWED THE IMAGES/STUDY AND I AGREE WITH THE FINDINGS AS STATED BY RESIDENT PHYSICIAN DR. RAELYNNE MACBETH.		0
					NO ACUTE OSSEOUS ABNORMALITY OF THE PELVIS, RIGHT HIP, OR RIGHT FEMUR.		

3.3 DATA ANALYSIS

So, our objective was to evaluate the diagnostic performance of AI algorithms in interpreting lower extremities musculoskeletal (MSK) X-rays and for this study various metrics are used:

- **Confusion Matrix:** A confusion matrix is a tool used to evaluate the performance of a machine learning model on test data. It displays the counts of correct and incorrect predictions made by the model. This matrix is particularly useful for assessing classification models, which are designed to assign a category label to each input instance.

- **Sensitivity:** Sensitivity, also known as recall or true positive rate, is a measure of a classification model's ability to correctly identify positive instances. It is defined as the ratio of true positives (correctly predicted positive cases) to the total number of actual positives (the sum of true positives and false negatives). A high sensitivity indicates that the model is effective at identifying positive cases.

- **Specificity:** Specificity, also known as the true negative rate, is a measure of a classification model's ability to correctly identify negative instances. It is defined as the ratio of true negatives (correctly predicted negative cases) to the total number of actual negatives (the sum of true negatives and false positives).

- **Precision:** Precision, also known as positive predictive value, is a measure of the accuracy of a classification model's positive predictions. It is defined as the ratio of true positives (correctly predicted positive cases) to the total number of predicted positives (the sum of true positives and false positives).
  Mathematically they are calculated by these formulas:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$= \frac{True\ Positive}{Total\ Predicted\ Positive}$$

- **Construction of Confusion Matrix:**

   i. First for developing the matrix we need two comparative values that are Predictive value which is provided by the AI and other is the actual value which is the Ground truth that is created according to the verdict of the radiologist.
   ii. GT of 4108 X-ray cases was created by reading the impressions and coded in the format of
      - 1 = Fracture present

- 0 = Fracture Absent
- 2 = Healed Fracture/Deformity
- 5 = Questionable Fractures

iii. GT was created according to the body part visualized under the scan and separate column for each body part was created to so in which ever bone the fracture is it will be marked in the concern bone column because in the single group there might be multiple fractures of different bones.

	Impression	Femur	Tibia	Ankle	Foot	hip	Fibula	patella	Should	Humer	Elbow	Radius	Ulna	Wrist_x	Hand_x	Scapho	arm	elbow	forearm
7	1. NO ACUTE FRACTURE OR MALALIGNMENT OF THE PELVIS OR BILATERAL HIPS, FEMURS OR KNEES. 2. DEGENERATIVE CHANGES AS ABOVE. I PERSONALLY REVIEWED THE IMAGES/STUDY AND I AGREE WITH THE FINDINGS AS STATED. THIS STUDY WAS INTERPRETED AT UNIVERSITY HOSPITALS CLEVELAND MEDICAL CENTER, CLEVELAND, OHIO.	0				0													
3	NO ACUTE FRACTURE OR MALALIGNMENT RIGHT FEMUR OR HIP. I PERSONALLY REVIEWED THE IMAGES/STUDY AND I AGREE WITH THE FINDINGS AS STATED ABOVE BY RESIDENT PHYSICIAN, DR. GREGORY R. LILLER. THE STUDY WAS INTERPRETED AT UNIVERSITY HOSPITALS CLEVELAND MEDICAL CENTER IN CLEVELAND OHIO.	0				0													

iv. For comparing the outputs of AI with the GT only the definite fracture or no fracture codes that are 1&0 are considered for this study

v. So, out of 4108 we will consider the 3767 cases with definite GTs

vi. Within these 3767 cases AI gives 2326 of 0 output and 1441 of 1 output

vii. Whereas according to the GT 2876 are 0 and 891 are 1

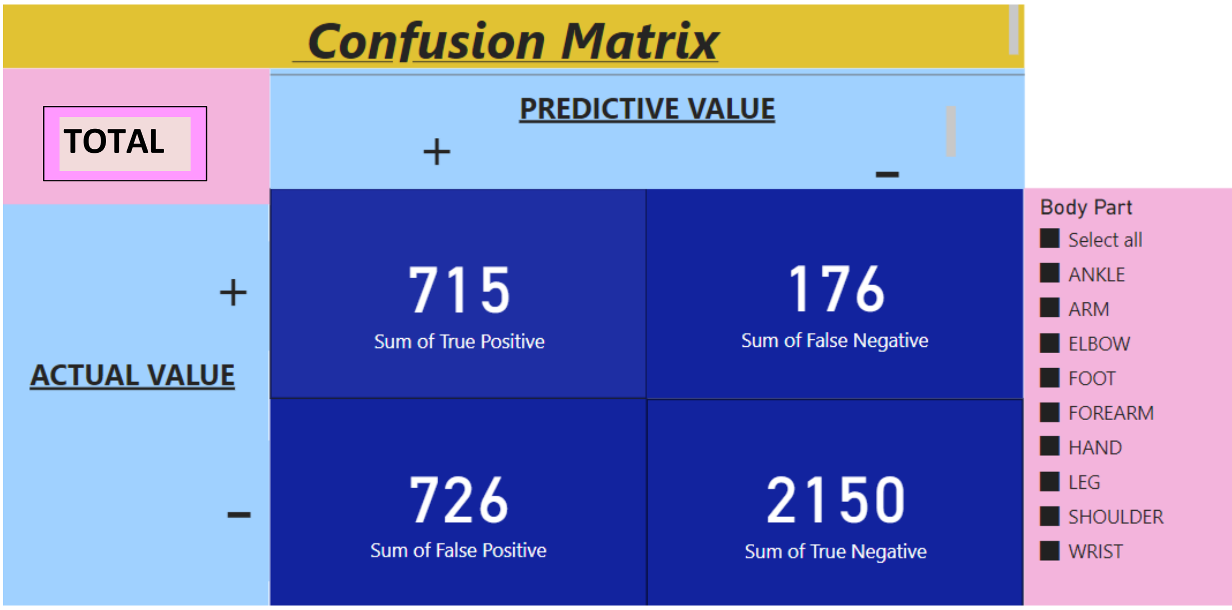
viii. Now we will calculate the
- True positives (both AI and GT =1)
- True negatives (both AI and GT=0)
- False positives (AI=1 and GT=0)
- False negatives (AI=0 and GT=1)

ix. Here, study is focused on lower bones, so the spreadsheet is divided into 4 parts Total (Comprising all the body parts), Leg, Ankle and Foot.

	A	B	C	D	E	F	G	H
1	Group	GT	AI	True Positi	True Negati	False Positi	False Negati	
2	SHOULDER	0	0	0	0	0	0	
3	FOREARM	0	0	0	0	0	0	
5	WRIST	0	0	0	0	0	0	
6	SHOULDER	0	0	0	0	0	0	
7	WRIST	0	0	0	0	0	0	
8	FOREARM	0	0	0	0	0	0	
9	HAND	0	1	0	0	0	0	
10	FOREARM	1	1	0	0	0	0	
11	WRIST	1	1	0	0	0	0	
12	HAND	0	0	0	0	0	0	
13	LEG	0	0	0	0	0	0	
14	LEG	0	0	0	0	0	0	
15	WRIST	1	1	0	0	0	0	
17	WRIST	0	0	0	0	0	0	
18	ELBOW	0	1	0	0	0	0	
19	ELBOW	0	0	0	0	0	0	
20	SHOULDER	0	0	0	0	0	0	
21	ARM	0	1	0	0	0	0	
22	WRIST	0	0	0	0	0	0	
23	HAND	0	0	0	0	0	0	
24	LEG	0	0	0	0	0	0	
25	ELBOW	0	0	0	0	0	0	
26	HAND	0	1	0	0	0	0	
27	WRIST	1	1	0	0	0	0	
28	FOREARM	1	1	0	0	0	0	
29	ELBOW	0	0	0	0	0	0	
30	WRIST	1	1	0	0	0	0	

x. Above Excel file is connected with Power BI to create the Dashboard to visualize the Confusion metrics of different body parts.

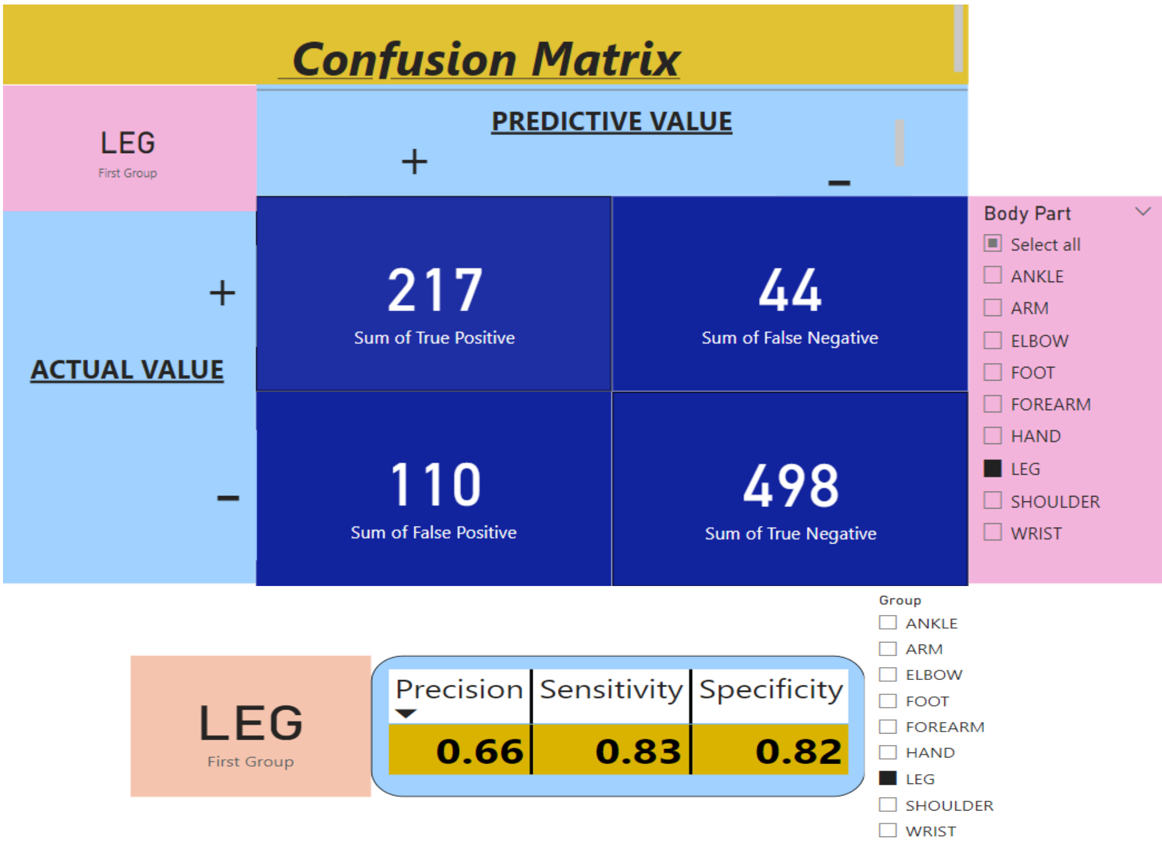
CONFUSION MATRIX FOR THE TOTAL BODY PARTS



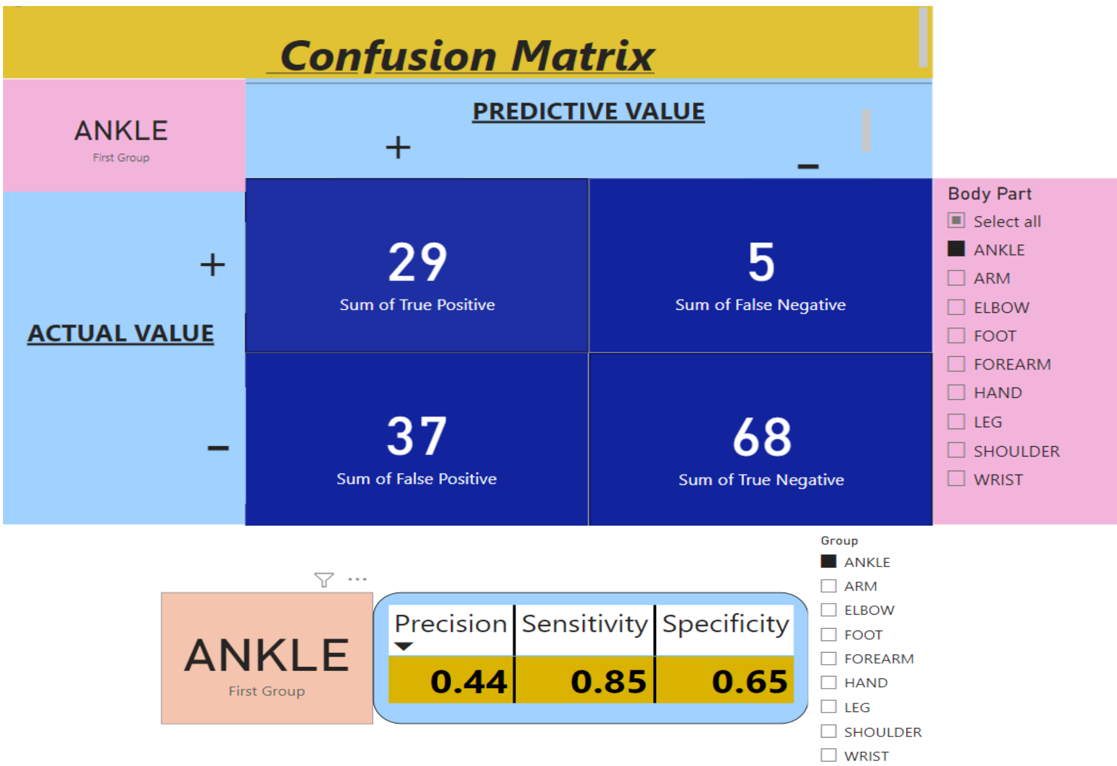
TOTAL

Precision	Sensitivity	Specificity
0.50	0.80	0.75

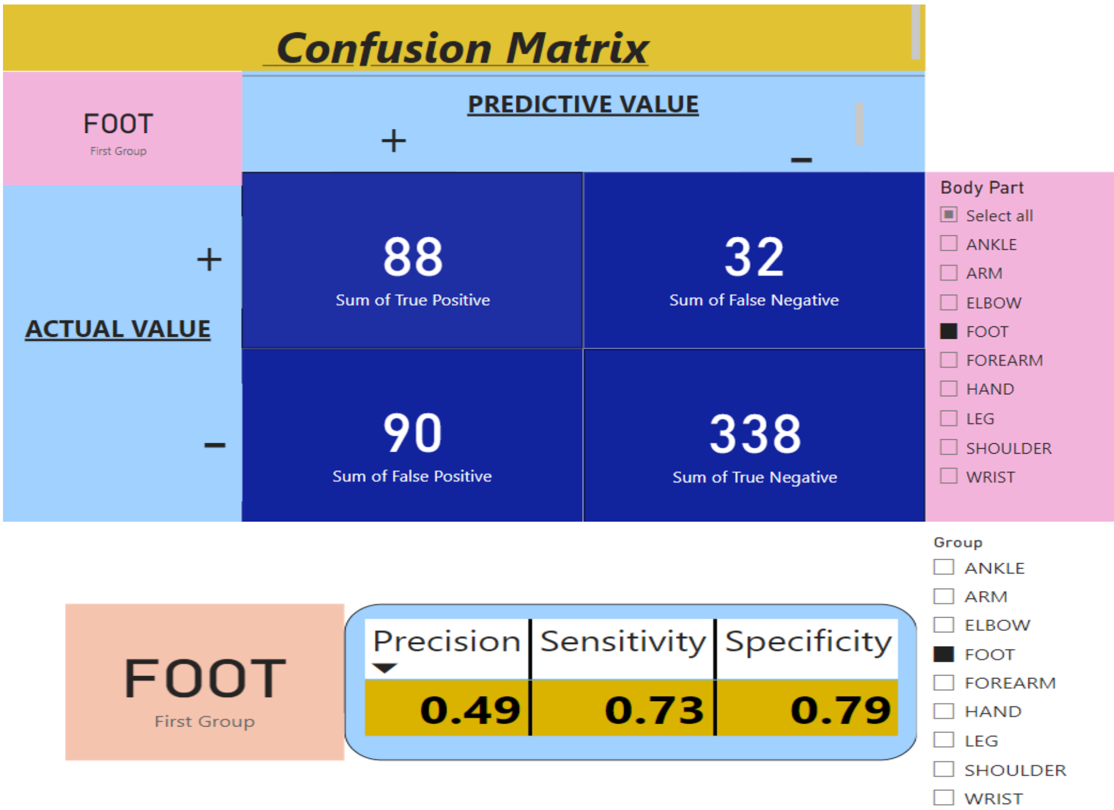
CONFUSION MATRIX FOR LEG



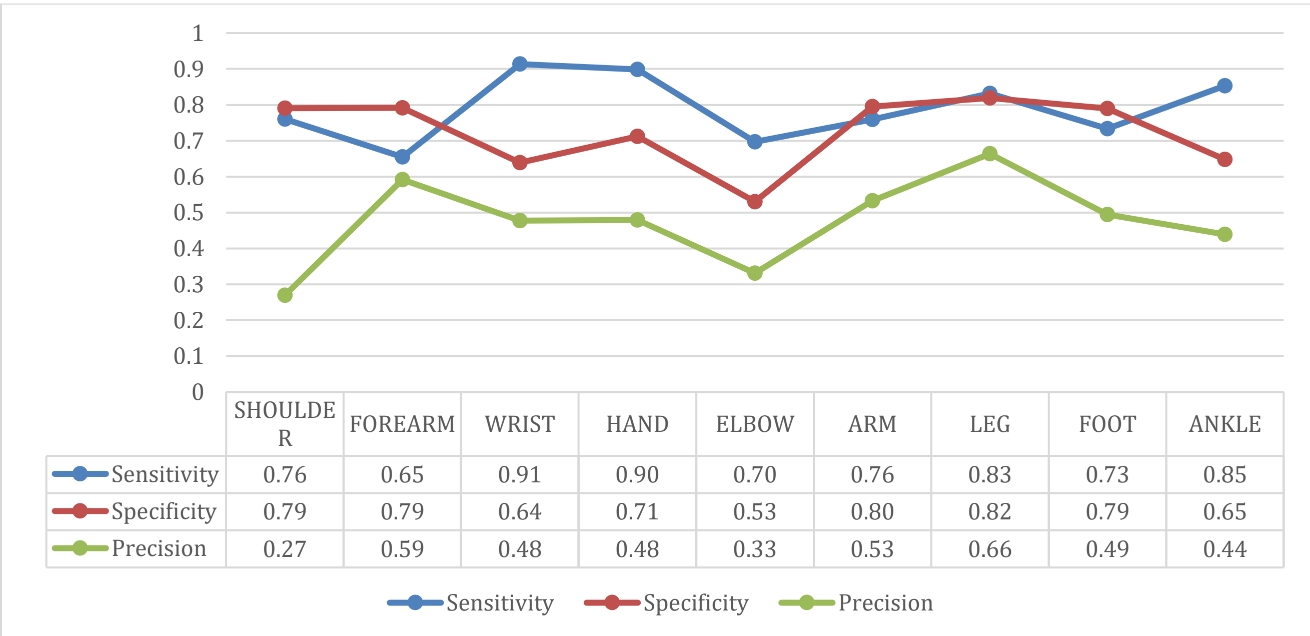
CONFUSION MATRIX OF ANKLE



CONFUSION MATRIX OF FOOT



Comparative Plot of AI performance metrics for different body parts:



	SHOULDER	FOREARM	WRIST	HAND	ELBOW	ARM	LEG	FOOT	ANKLE
Sensitivity	0.76	0.65	0.91	0.90	0.70	0.76	0.83	0.73	0.85
Specificity	0.79	0.79	0.64	0.71	0.53	0.80	0.82	0.79	0.65
Precision	0.27	0.59	0.48	0.48	0.33	0.53	0.66	0.49	0.44

3.3 DATA INTERPRETATIONS

The evaluation of AI algorithms in interpreting musculoskeletal (MSK) X-rays focused on the lower extremities, including the leg, ankle, and foot. The study utilized confusion matrices to compare AI findings with the ground truth (GT) established by radiologists. Key performance metrics such as sensitivity, specificity, and precision were calculated for each body part to assess the diagnostic accuracy of the AI.

AI performance for detecting fractures across all the parts:

- **Sensitivity (Recall/True Positive Rate):** Sensitivity measures the AI's ability to correctly identify fractures (positive cases). For the total body parts, a sensitivity of 0.80 indicates that the AI correctly identified 80% of the actual fractures. This high sensitivity suggests the AI is effective at detecting fractures, reducing the risk of missed diagnoses which is crucial for patient safety.
- **Specificity (True Negative Rate):** Specificity indicates the AI's ability to correctly identify non-fractures (negative cases). The total specificity of 0.75 shows that the AI correctly identified 75% of the actual non-fracture cases. This reflects a moderate rate of false positives, meaning some patients without fractures might be incorrectly flagged, leading to unnecessary further testing.
- **Precision (Positive Predictive Value):** Precision measures the accuracy of the AI's positive predictions. With a precision of 0.50 for total body parts, the AI's correct fracture predictions are only half of all the predicted fractures. This lower precision suggests a significant number of false positives, indicating that while the AI is sensitive, it also tends to over-predict fractures.

The major concern in the predictive ability of the AI algorithm is the number of True positive and False positives are quite similar in fact False positives are slightly higher than True positives which indicates the AI is overcalling fractures and this is the ultimate reason of moderate Specificity and low precision.

AI performance for detecting fractures of Leg:

- **Sensitivity (0.83):** The AI correctly identifies 83% of actual leg fractures, indicating a high true positive rate. This high sensitivity is crucial for ensuring that most fractures are detected, minimizing the risk of missed diagnoses.
- **Specificity (0.82):** The AI correctly identifies 82% of non-fractured leg cases. This high specificity shows that the AI is effective at ruling out non-fracture cases, reducing unnecessary concern or treatment.
- **Precision (0.66):** The AI's positive leg fracture predictions are correct 66% of the time. While this is a decent precision rate, it indicates some false positives, meaning a portion of the predicted fractures are not actually fractures.

Risks:

- The presence of moderate false positives can lead to unnecessary anxiety for patients and additional diagnostic procedures, potentially increasing healthcare costs and patient burden.

AI performance for detecting fractures of Ankle:

- **Sensitivity (0.85):** The AI detects 85% of actual ankle fractures, showing excellent performance in identifying fractures and ensuring that most cases are caught.
- **Specificity (0.65):** The AI correctly identifies 65% of non-fractured ankle cases. This moderate specificity indicates some false positives, which could lead to unnecessary follow-up tests.
- **Precision (0.44):** The precision rate of 44% indicates that less than half of the AI's predicted ankle fractures are correct. This relatively low precision suggests that the AI tends to over-predict fractures in the ankle, resulting in many false positives.

Risks:

- The low precision suggests a substantial number of false positives, leading to potential over-diagnosis and unnecessary interventions. This could strain medical resources and result in patient distress.

- The moderate specificity might contribute to mismanagement in clinical settings, as the AI could mislead clinicians into suspecting fractures where there are none.

AI performance for detecting fractures of Foot:

- **Sensitivity (0.73):** The AI accurately identifies 73% of actual foot fractures, which is reasonably good but indicates room for improvement in catching more true positive cases.
- **Specificity (0.79):** The AI correctly identifies 79% of non-fractured foot cases, showing strong performance in ruling out non-fractured cases and reducing unnecessary treatment.
- **Precision (0.49):** The AI's foot fracture predictions are correct 49% of the time. This indicates a significant number of false positives, suggesting that the AI's predictions in the foot area are less reliable and need refinement.

Risks:

- The lower sensitivity compared to other areas suggests a risk of missing some foot fractures, which could lead to delayed or incorrect treatment.
- The precision indicates many false positives, causing potential over-treatment and unnecessary follow-up tests.
- The moderate performance across all metrics for the foot could create diagnostic uncertainty, potentially affecting clinician trust in the AI system for this body part

Comparative Insights:

- **Leg vs. Ankle:** The AI performs better for the leg than the ankle in terms of specificity and precision. The higher specificity for the leg suggests fewer false positives, and better precision indicates more accurate predictions.
- **Leg vs. Foot:** The leg also shows better performance metrics than the foot, particularly in sensitivity and precision, making it the most reliable body part for the AI diagnosis among the three.
- **Ankle vs. Foot:** The ankle demonstrates higher sensitivity, but lower specificity and precision compared to the foot. This trade-off implies that while the ankle model is better at detecting fractures, it also generates more false positives.

The AI model was developed for the fracture detection and the sensitivities is the proof of successful detection of the fractures but the model is providing higher false positives as well as the low precision too which questions the efficiency and the reliability on the usage of AI in clinical practices. For this again I referred to the cases where there is overcalling of fracture by the AI and found some of the cases which might be the possible reason for the:

- **Curvilinear densities**
- **Effusion**
- **Degenerative changes**
- **Deformities**
- **Ossification**
- **Osteotomy**
- **Dislocations**
- **Arthrodesis**
- **Osteoarthritic changes**
- **Syndesmoses**
- **Coalition**

AI Misclassifications:

1. Visual Similarities in Imaging

Many of the conditions misclassified by the AI have imaging characteristics that can appear like fractures:

- Curvilinear Densities: These can mimic the appearance of fracture lines due to their shape and contrast in X-rays.
- Effusion: Joint effusions might obscure the joint space or margins, making it challenging for the AI to distinguish between fluid accumulation and fracture lines.
- Degenerative Changes and Osteoarthritic Changes: The irregularities and bone spurs associated with degenerative changes can be mistaken for fracture fragments.
- Ossification and Osteotomy: The presence of new bone formation or surgical changes might present as irregular bone margins, similar to fracture lines.
- Dislocations and Deformities: These can cause anatomical distortions that the AI might

misinterpret as fractures.

- Arthrodesis and Syndesmoses: These surgical and anatomical fusions might appear as abnormal bone continuity, confusing the AI.
- Coalition: Bone fusions in coalition can mimic the continuity disruptions seen in fractures.

2. Training Data Limitations

The AI model's training data may not have adequately differentiated between fractures and the listed clinical conditions. If the training dataset included insufficient examples of these conditions or did not explicitly label them as non-fractures, the AI might not have learned to distinguish between them effectively.

3. Feature Overlap

The features used by the AI to detect fractures, such as edge detection, density gradients, and discontinuities in bone structure, might also be present in the other conditions. This feature overlap can lead to false positives where the AI incorrectly identifies these features as indicative of fractures.

4. Algorithm Complexity and Interpretability

Complex AI models, particularly deep learning models, can sometimes develop decision-making processes that are not easily interpretable. This black-box nature makes it difficult to pinpoint the exact reasons for misclassifications without extensive analysis and refinement of the model's parameters and training process.

Scope of Improvement:

1. **Enhanced Training Data:**
    - Incorporating a larger and more diverse set of X-ray images, including those with the abovementioned conditions, can help the AI learn to distinguish between fractures and other clinical abnormalities.
    - Annotating these images with detailed labels indicating the presence of curvilinear densities, effusion, degenerative changes, etc., can aid the AI in differentiating these from actual fractures.

2. **Regular Model Updates:**
   - Continuously updating the AI model with new data and feedback from radiologists can help it adapt to new patterns and improve its classification accuracy over time.
   - Conducting regular performance reviews and retraining the model to address specific weaknesses identified through false positive analysis.

3. **Collaborative Diagnosis:**
   - Using the AI model as an assistive tool rather than a standalone diagnostic system can help mitigate risks. Clinicians can use AI predictions as a second opinion, verifying its findings with their expertise.

SECONDARY OBJECTIVES

3.4 DATA COMPILATION

So, for further analysis, the data of 663 MSK X-ray reports interpreted by the junior radiologist and the senior radiologist was provided. Out of these 663 the junior radiologist is assisted by the AI in 329 cases where as in 334 cases no AI assistance was being used. My study was centric to the lower extremities so the proceedings of the data analysis will be according to the concordance and discordance of the interpretations of the lower body parts.

This data was also provided in the CSV format so the data was also converted into columns by removing the delimiter.

Now GT was calculated for the interpretations of the scans and in this case for the same accession number two interpretations were there one from the junior radiologist and one from the senior radiologist. The GT was prepared for both the interpretations. The GT was coded in the same way i.e.:

1= Fracture Present

0= Fracture Absent

2=Healed Fractures and Deformities

5= Questionable Fractures

The GT was prepared part specific and within the same column coding was done to represent presence or absence of fracture in different bones –

(H, W, R, U, HE, C, S, HIP, FE, TI, FI, A, FO) = (Hand, Wrist, Radial, Ulna, Humerus, Clavicle, Scapula, Hip, Femur, Tibia, Fibula, Ankle, and Foot).

After preparing the GT the data was separated into two sheets on sheet was with AI assistance and one was without AI.

	A	B	C	D
1	MRN	JR(H,W,R,U,He,C,S,Hip,F,P,T,Fi,A,F	SR(H,W,R,U,He,C,S,Hip,F,P,T,Fi,A,I	Disconcordance
2	8381423	0,0,1,1,0,0,0,0,0,0,0,0,0,0	0,0,1,1,0,0,0,0,0,0,0,0,0,0	Concordant
3	1332680	0,0,0,0,0,0,0,0,0,0,1,0,0,0	0,0,0,0,0,0,0,0,0,0,1,0,0,0	Concordant
4	1293856	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
5	1804778	0,0,0,0,0,0,0,0,0,0,0,1,0,1	0,0,0,0,0,0,0,0,0,0,0,1,0,1	Concordant
6	7755669	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
7	1082338	0,0,0,0,0,0,0,0,0,0,0,0,0,1	0,0,0,0,0,0,0,0,0,0,0,0,0,1	Concordant
8	1096135	0,0,0,0,0,0,0,0,0,0,1,0,0,0	0,0,0,0,0,0,0,0,0,0,1,0,0,0	Concordant
9	1154494	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
10	1588870	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
11	1981302	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
12	289936	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
13	753340	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
14	429263	0,0,1,1,0,0,0,0,0,0,0,0,0,0	0,0,1,1,0,0,0,0,0,0,0,0,0,0	Concordant
15	7586103	5,0,0,0,0,0,0,0,0,0,0,0,0,0	1,0,0,0,0,0,0,0,0,0,0,0,0,0	Disconcordant
16	1998942	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
17	1219709	5,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Disconcordant
18	2319555	1,0,0,0,0,0,0,0,0,0,0,0,0,0	1,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
19	1178472	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,5,0,0,0	Disconcordant
20	1686972	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
21	828833	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
22	1178472	0,0,0,0,0,0,0,0,0,0,5,0,0,0	0,0,0,0,0,0,0,0,0,0,5,0,0,0	Concordant
23	8012630	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
24	3337081	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,0,0	Concordant
25	1079993	0,0,0,0,0,0,0,0,0,0,1,0,5,0	0,0,0,0,0,0,0,0,0,0,1,0,5,0	Concordant
26	8260882	0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,0,0,0,0,0,0,0,5,0	Disconcordant

3.5 DATA ANALYSIS

Now the GTs for both the interpretations of the Junior and Senior radiologists are compared first in the cases of 329 with AI assistance and then 334 without AI assistance. The study focuses on the lower extremities that the GTs concordance and discordance will be compared according to the presence or absence of Fractures in the bones like: Hip, Femur, Tibia, Fibula, Ankle and Foot.

VARIABLES IN THE STUDY:

- **Independent Variable (IV):**
  - **AI Assistance:** This variable indicates whether the diagnosis was assisted by AI or not.
    - AI-assisted
    - Non-AI-assisted

- **Dependent Variables (DVs):**
  - **Concordance:** This variable measures the number of cases where the AI's interpretation agreed with the JR and SR.
  - **Discordance:** This variable measures the number of cases where the AI's interpretation did not agree with the JR and SR.

- **Categorical Variables:**
  - **Diagnosis Outcome:** This variable categorizes the outcome of the diagnosis.
    - Concordant
    - Discordant

With AI assistance cases:

- Total = 329 cases
- Concordance = 320 cases
- Percentage (Concordance) = 97.3%
- Discordance = 9 cases
- Percentage (Discordance)=2.7%

Without AI assistance cases:

- Total = 334
- Concordance = 319 cases
- Percentage (Concordance) = 95.5%
- Discordance = 15 cases
- Percentage (Discordance) = 4.5%

HYPOTHESIS FORMULATION

- **Null Hypothesis:** There is no statistically significant difference in concordance and discordance of AI interpretations by JR and SR between AI assisted and non- AI assisted cases.

- **Alternative Hypothesis:** There is a statistically significant difference in concordance and discordance of AI interpretations by JR and SR between AI assisted and non- AI assisted cases.

HYPOTHESIS TESTING

So, for obtaining a concrete proof for the rejecting and accepting the null hypothesis or alternate hypothesis some statistical hypothesis tests are used. For this data we will compare the discordance and concordance in the reports with and without AI assistance. This a type of categorical data where there is a binary output of presence of discordance YES or NO, if No then case of concordance and this is being compared in both the groups with and without AI.
In this the most suitable test is Chi- square test i.e. A chi-square test is a statistical method used to compare observed data with expected data to determine if any differences are due to chance or if there is a relationship between the variables being studied.

- First 2x2 Contingency Table was prepared:

2X2 Table	Concordant	Discordant	Row Total
AI-assisted	320	9	329
Non-AI-assisted	319	15	334

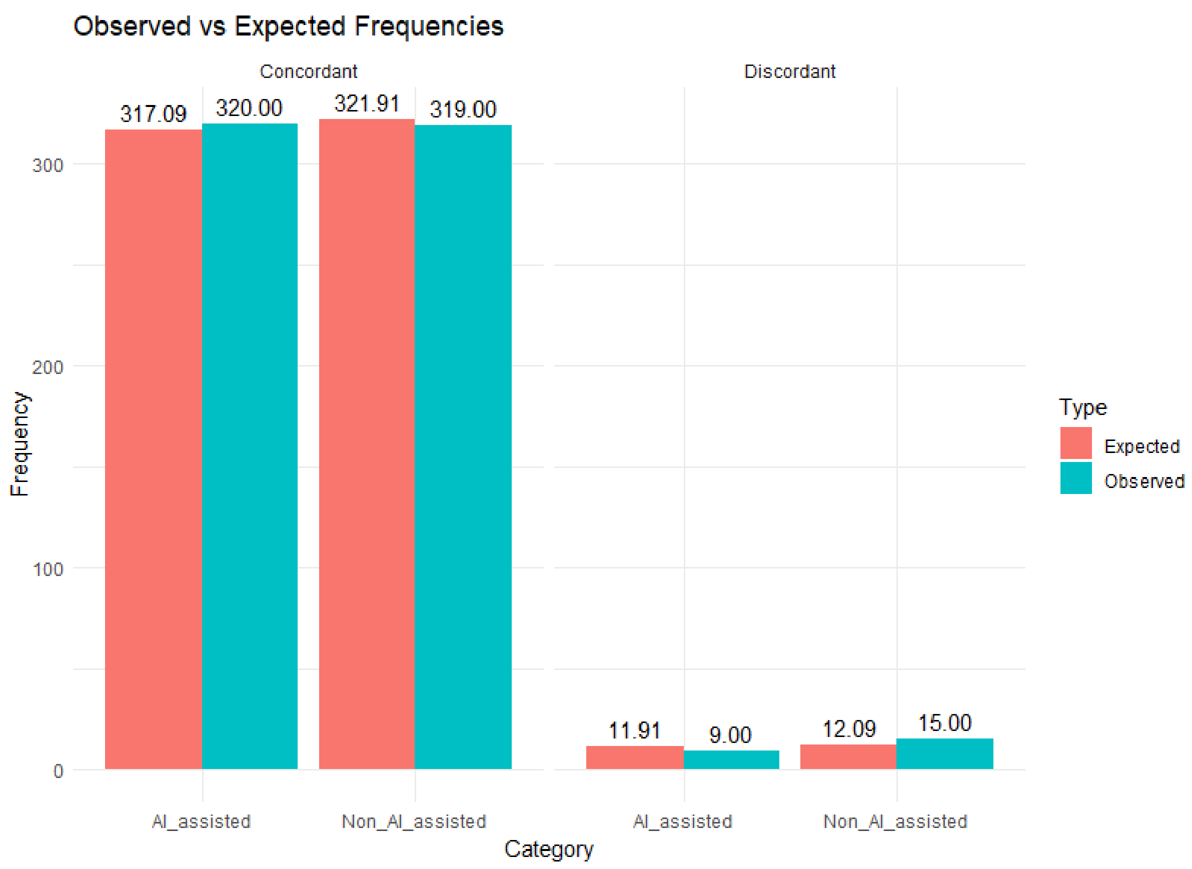
Column Total	639	24	663

- Now calculation of Expected Frequencies by the formula:

$$Expected = \frac{(row\ total\ \times column\ total)}{overall\ total}$$

- 2x2 Contingency Table for Expected Frequencies

2X2 Table	Concordant	Discordant	Row Total
AI-assisted	317.1	11.9	329
Non-AI-assisted	321.9	12.1	334
Column Total	639	24	663



- **R programming language has been used for applying the Chi square test:**

```r
install.packages("dplyr")
library(dplyr)


observed <- matrix(c(320, 9, 319, 15), nrow = 2, byrow = TRUE)
colnames(observed) <- c("Concordant", "Discordant")
rownames(observed) <- c("AI_assisted", "Non_AI_assisted")


total_cases <- sum(observed)
row_totals <- rowSums(observed)
col_totals <- colSums(observed)
expected <- outer(row_totals, col_totals) / total_cases


chi_square_test <- chisq.test(observed, expected, correct = FALSE)

print(chi_square_test)

```

19:1 (Top Level)

Console Terminal Background Jobs

R 4.3.1 · ~/

```
>
>
> total_cases <- sum(observed)
> row_totals <- rowSums(observed)
> col_totals <- colSums(observed)
> expected <- outer(row_totals, col_totals) / total_cases
>
>
> chi_square_test <- chisq.test(observed, expected, correct = FALSE)
>
> print(chi_square_test)

        Pearson's Chi-squared test

data:  observed
X-squared = 1.4639, df = 1, p-value = 0.2263
```

So, Chi Square Value is 1.4639 which is the test statistics and p value corresponding to it is 0.2263

3.6 DATA INTERPRETATION

Confidence Interval = 95%

Degree of Freedom= 1

Chi square value (test statistics) = 1.4639

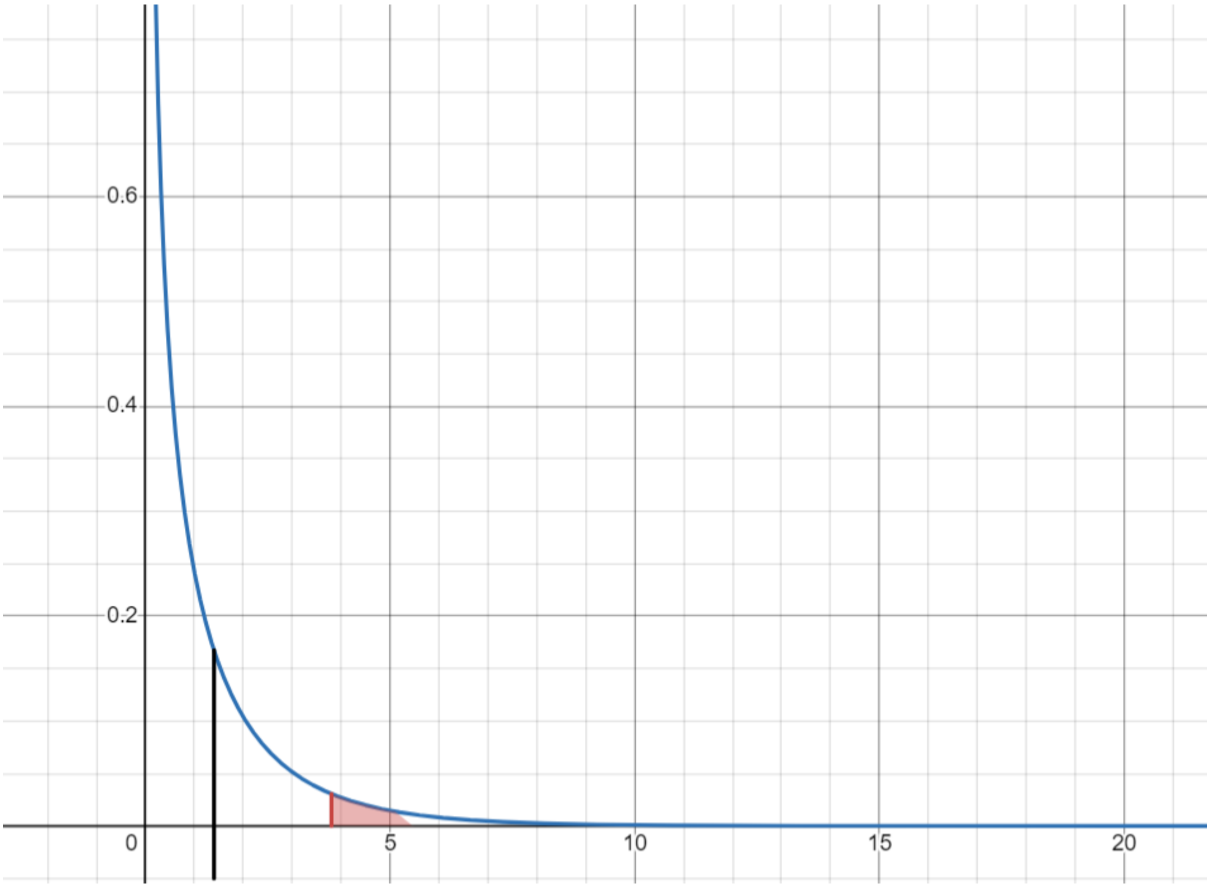
p value = 0.2263

At the CI 95% the p value that is the level of significance = 0.05

Corresponding to it, the critical value that is the tabled value of Chi Square= 3.84

Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09



The Chi square probability distribution graph shows the red highlighted area is the rejection area that is the 0.05 p value initiated from the critical value 3.84 whereas the black line preceding to the rejection area shows the test statistics value of 1.46.

- So, the Test statistics value < Critical value that's why it fell in the acceptance region
- The p value (0.2263) > p value (0.05)

Conclusion Based on Hypothesis Testing

- Fail to reject the null hypothesis.
- The data does not provide sufficient evidence to conclude that there is a statistically significant difference in the concordance and discordance rates between AI-assisted and non-AI-assisted cases. This suggests that the use of AI assistance does not significantly affect the agreement (concordance) or disagreement (discordance) of diagnostic interpretations by JR and SR radiologists.

Analysis and Insights

1. **Concordance Rates:**
   - **AI-Assisted Cases:** 320 out of 329 cases (97.3%) were concordant.
   - **Non-AI-Assisted Cases:** 319 out of 334 cases (95.5%) were concordant.
2. **Discordance Rates:**
   - **AI-Assisted Cases:** 9 out of 329 cases (2.7%) were discordant.
   - **Non-AI-Assisted Cases:** 15 out of 334 cases (4.5%) were discordant.

Insights:

- **Consistency in Performance:** The high concordance rates for both AI-assisted and non-AI-assisted cases indicate that AI assistance does not significantly alter the overall performance of radiologists. Both methods yield high accuracy rates, suggesting that AI can be a reliable support tool without introducing significant errors.
- **Reliability of AI:** The similar concordance and discordance rates suggest that the AI system's interpretations are generally in agreement with those of human radiologists. This represents the potential ability of AI in diagnostic settings, where it can provide a second opinion or assist in decision-making.
- **Focus on Improvement:** Despite the lack of a significant difference, the slight variations in discordance rates highlight that if the model is trained and tested continuously it has the potential to give positive results.

- **Discordance Rate**: The slightly higher discordance in non-AI-assisted cases (4.5%) compared to AI-assisted cases (2.7%) implies that AI assistance may reduce some interpretation errors by providing additional data points or highlighting areas of concern that radiologists might miss.

But the lack of statistically significant difference is the cause that these-insight might not be concrete enough to be claimed in a clinical setting

PRACTICAL APPLICATION OF AI IN CLINICAL SETTINGS AND ITS POTENTIAL BENEFITS AND LIMITATIONS FOR PATIENTS

Benefits of AI in Clinical Settings

1. **Enhanced Diagnostic Accuracy:**
   - **High Concordance:** The high concordance rates suggest that AI can help radiologists make more accurate diagnoses by providing a second opinion and highlighting areas of concern that may be missed.
   - **Reduced Discordance:** Lower discordance rates with AI assistance indicate that AI can help minimize misdiagnoses, potentially leading to better patient outcomes.
2. **Increased Efficiency:**
   - **Quick Analysis:** AI systems can quickly analyze medical images, allowing radiologists to focus more on complex cases and reduce their workload.
   - **Support for Junior Radiologists:** AI assistance can be particularly beneficial for junior radiologists, providing them with reliable support and increasing their diagnostic confidence.

Limitations of AI in Clinical Settings

1. **False Positives:**
   - **Lower Precision:** The study indicates that the AI model has a lower precision, especially for the ankle and foot, leading to more false positive results. This can result in unnecessary further testing and anxiety for patients.
   - **Misclassification of Other Conditions:** The AI model sometimes detects conditions like curvilinear densities, effusion, degenerative changes, etc., as fractures, contributing to false positives.

2. **Dependency on Quality of Data:**
    - **Training Data Limitations:** The performance of AI models is highly dependent on the quality and diversity of the training data. If the training data does not adequately represent the range of conditions seen in clinical practice, the AI model may not perform well in all scenarios.
3. **Clinical Integration Challenges:**
    - **Adoption Barriers:** Integrating AI into clinical workflows can be challenging due to resistance from medical staff, the need for extensive training, and the requirement for robust IT infrastructure.
    - **Ethical and Legal Considerations:** There are concerns about liability in cases where AI-assisted diagnoses lead to incorrect treatment decisions.

Section-4: RECOMMENDATION

- Might provide statistical results if the research will be done with a bigger sample size which will satisfy the effect size
- AI is needed to be trained with more varied and clinical cases for avoiding misclassification
- AI can be used only as an assistive agent rather than loading full reliability over its interpretations.
- Questionable scenarios might be there in which the use of AI outputs was not considered by the Junior radiologist due to resistance in change within the workflow. This can be overlooked by training and workshop of clinical professionals.

CONCLUSION

The study demonstrates the potential of AI to enhance diagnostic accuracy by high sensitivity and specificity in clinical settings. The high concordance rates with AI assistance and the reduced discordance rates suggest that AI can serve as a valuable tool for radiologists, particularly in supporting junior staff and handling routine cases efficiently. However, the limitations highlighted by the study, such as false positives and the dependency on high-quality training data, underscore the need for continuous improvement and careful integration of AI systems into clinical practice. Balancing these benefits and limitations is crucial for leveraging AI's full potential while ensuring patient safety and maintaining trust in medical diagnostics.

SECTION-5: REFERENCES

1. Lindsey R, Daluiski A, Chopra S, et al. Artificial Intelligence in Fracture Detection: A Systematic Review. 2021.
2. Brady AP. The discrepancy rate between preliminary and final reports of emergency radiology studies: A performance indicator and quality improvement method. 2012.

4. Patel B, Taylor C, Khong PL, et al. Artificial Intelligence in Radiology: Current Applications and Future Directions. 2021.
5. Parakh A, Kortman H, Hermans M, et al. The Role of Artificial Intelligence in Emergency Radiology: What the Radiologist Needs to Know. 2021.
6. Scheinfeld MH, Rahman S, Fishman EK, et al. Artificial Intelligence in Radiology: Progress, Pitfalls, and Potential. 2023.

Swastik Mitra ST report

ORIGINALITY REPORT

7%	5%	3%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Student Paper	1%
2	www.coursehero.com Internet Source	<1%
3	Submitted to Angel Group Student Paper	<1%
4	mbaskills.in Internet Source	<1%
5	robots.net Internet Source	<1%
6	www.aal-persona.org Internet Source	<1%
7	Barazanji, Omar. "Automatic Actigraphy and Polysomnography Sleep Scoring Using Deep Learning", Auburn University, 2023 Publication	<1%
8	Submitted to National & Kapodistrian University of Athens Student Paper	<1%