EXPLORING ARTIFICIAL INTELLIGENCE BASED INTERPRETATION OF MEDICAL IMAGES IN RADIOLOGICAL WORKFLOWS IN ZAMBIA

By

Gabriel Malaizyo Muzumala

A Thesis submitted to the University of Zambia, in fulfilment of the requirements for the award of the degree of Master of Science in Computer Science

The University of Zambia

December, 2025

Copyright Declaration

© 2024 by Malaizyo Gabriel Muzumala. All rights reserved

Declaration

I, Malaizyo Gabriel Muzumala , hereby affirm that this dissertation is solely my original creation and has not been presented for consideration at any other college, institution, or university apart from the University of Zambia.

Name: …….

Sign: ……

Date: ……

Abstract

The manual nature of radiological workflows is a major issue affecting patient outcomes and the quality of healthcare provision in Zambia with only 9 radiologists in 2021. Artificial intelligence (AI) solutions are increasingly becoming mainstream. However the implementation and use of these solutions in low resource settings has been very slow. This study aimed to: (1) empirically characterize the workflow challenges regarding workload and effectiveness in Zambian radiology; (2) design and implement a locally adaptable, semi-automated AI solution for pneumonia detection; and (3) evaluate the impact of this solution on practitioner workload and user acceptance. Guided by the Cultural-Historical Activity Theory (CHAT) framework, a multi-phase mixed-methods design was employed. Phase I utilized a cross-sectional survey (n=100) and semi-structured interviews to assess baseline constraints. Phase II involved the development of binary classification (VGG16) and object detection (Mask R-CNN) models integrated into a web-based DICOM viewer. Phase III consisted of a controlled, within-subjects experiment (N=47) comparing manual versus AI-assisted interpretation using the NASA Task Load Index (NASA-TLX) and Technology Acceptance Model (TAM 2). Contextual inquiry revealed significant systemic breakdowns, with 74.5% of clinicians reporting reduced effectiveness due to missing clinical history and a heavy reliance on non-dedicated viewing devices. The developed AI models achieved a high Recall of 79%, prioritizing sensitivity to serve as a safety net for screening. Experimental results indicated a reduction in perceived workload, specifically in Mental Demand (Raw Score Δ-0.68), and high user acceptance, with a mean Perceived Usefulness of 5.7/7. The study concludes that human-in-the-loop AI effectively mediates the tension between high patient volumes and limited expertise by functioning as a cognitive scaffold for generalist doctors. It is recommended that health stakeholders adopt AI-enabled task shifting as a national strategy, provided it is accompanied by investments in digital infrastructure and mandatory user competence training to ensure safe deployment.

Acknowledgements

First and foremost I would like to thank my supervisors.

Dr Lighton Phiri for being so present and serving as a mentor not just in my research journey but in my life journey as well. Giving me advice and skills that have changed the way I work and given me great opportunities through how they've made me stand out.

Dr Mayumbo Nyirenda for taking me under his wing in educational and professional life, bringing my name up and recommending me for opportunities including the scholarship through which I did this masters. Being kind and patient with me through my delays and going out of his way to assist in my journey.

Dr Ernest Zulu for being my guide in the medical environment, for sparing time to interact with tools and share the opinions on what we're working on and researching. For also the moments we've shared interesting life stories.

I would also like to thank the Entire Enterprise Medical Imaging in Zambia team, thanks so much for the assistance and presence. Dr Peter Chibuta, Andrew Shawa and Elijah Chileshe have moved from colleagues to being friends and I'm so grateful for all the sacrifice that has gone into making this work.

I would also like to thank the University Teaching Hospitals Department of Radiology in the Adult Hospital in which I practically became a member. I was graciously accepted by the radiologists and others and it greatly assisted with the research we're conducting.

Finally, I would like to thank the people closest to me, my Fiance Dr Thokozile Kaluba and family, thank you all so much for the support throughout this journey.

Table of Contents

Abstract IV
Acknowledgements V
Table of Contents VI
List of Tables XI
List of Figures XII
List of Abbreviations in Alphabetical Order XIII
Chapter 1. Introduction 1
1.1 Opening Context and Background 1
1.2 Problem Statement 2
1.3 Research Gap and Justification 3
1.4 Research Aim 3
1.5 Research Objectives and Questions 4
1.5.1 Research Objectives 4
1.5.2 Research Questions 4
1.6 Significance of the Study 4
1.7 Thesis Outline 5
1.8 Chapter Summary 5
Chapter 2. Related Work 6
2.1 The Clinical Context: Radiology in Resource-Limited Settings 6
2.1.1 Shortage of radiologists 7
2.1.2 Infrastructure and resource constraints 8
2.1.3 Challenges with Imaging data 8
2.2 Comparative Studies and Lessons from Similar Contexts 9
2.2.1 AI in Medical Imaging in Other LMICs 9
2.2.2 Successes and Failures 10
2.3 Technical State-of-the-Art: Deep Learning in Medical Imaging 11
2.3.1 Deep Learning Classification 11
2.3.2 Object Detection and Segmentation: Review Mask R-CNN 12
2.3.3 AI in the Global South: Challenges of Deployment 14
2.4 Human-AI Interaction and Evaluation Frameworks 15
2.4.1 Metrics for Clinical AI: The Accuracy Paradox 15
2.4.2 Measuring Cognitive Load and Acceptance 16
2.5 Regulatory and Ethical Considerations 17
2.5.1 Data Privacy and Patient Consent in AI Diagnostics 17
2.5.2 Ethical Concerns Around AI Decision-Making in Healthcare 18
2.5.3 Regulations Around AI in Healthcare 18

2.6 Datasets and Benchmarking in Medical Imaging AI 18
2.6.1 Publicly Available Datasets 18
2.6.2 Limitations of Existing Datasets for Global South Representation 19
2.6.3 Challenges in Acquiring Annotated Data 19
2.7. Human-AI Collaboration 20
2.7.1 Radiologist-in-the-Loop Systems 20
2.7.2 AI as a Triage Tool vs. Decision Support 20
2.8 Sustainability and Scalability of AI Solutions 20
2.8.1 Cost and Maintenance of AI Systems in Resource-Limited Settings 20
2.8.2 Training and Local Capacity Building 20
2.8.3 Long-Term Adoption and Support Challenges 21
2.9 Chapter Summary 21
Chapter 3. Research Methodology 23
3.1 Research Philosophy 23
3.1.1 Philosophical Stance 23
3.2 Theoretical Framework: Cultural-Historical Activity Theory (CHAT) 24
3.2.1 Core Components of the CHAT Model in this Study 25
3.2.2 CHAT as a Lens for Understanding Integration Dynamics 25
3.2.3 Why CHAT is Relevant for the Zambian Context 26
3.3 Research Design 26
3.3.1 The CRISP-DM Framework in this Study 27
3.3.2 Specific Research Strategies per Objective 28
3.3.3 Time Horizon 29
3.4 Phase I: Business Understanding (Contextual Inquiry) 29
3.4.1 Problem Scoping and Target Pathology 29
3.4.2 Study Settings and Target Population 30
3.4.3 Quantitative Data Collection: Clinician Survey 30
3.4.4 Qualitative Data Collection: Semi-Structured Interviews 32
3.4.5 Data Analysis Strategy for Phase I 33
3.5 Phase II & III: Data Understanding and Preparation Strategy 35
3.5.1 Dataset Selection Criteria 36
3.5.2 Strategy for Data Preparation and Augmentation 36
3.6 Phase IV: Modeling Strategy 38
3.6.1 Rationale for Transfer Learning (VGG16) 38
3.6.2 Rationale for Object Detection 38
3.6.3 Training Protocols 39
3.7 Phase V: Deployment Strategy (Iterative Design) 40
3.7.1 The User-Centered Design Process 40
3.7.2 Focus Group Methodology 40

3.8 Phase VI: Evaluation Methodology 42
3.8.1 Experimental Design 43
3.8.2 Experimental Setting and Procedure 44
3.8.3 Instruments 45
3.8.4 Data Analysis Procedure 46
3.9. Ethical Considerations 48
Chapter 4: System Design and Implementation 49
4.1 User-Driven Design Requirements 50
4.2 System Architecture 51
4.2.1 Architectural Components and Interaction Flow 52
4.2.2 Deployment Environment 54
4.3 Data Pipeline Implementation 55
4.3.1 Image Preprocessing and Normalization Logic 55
4.3.2 The Data Augmentation Heuristic 56
4.4 AI Model Implementation 58
4.4.1 Pneumonia Classification: VGG16 Architecture 58
4.4.2 Detection Model: Mask R-CNN Specifications 59
4.3.3 Multi-Stage Optimization Strategy 61
4.5 Application Interface Implementation 62
4.5.1 Web Interface Architecture 63
4.5.2 User Interface Evolution 65
4.6 Server-Side Logic and API Implementation 67
4.6.1 API Endpoint Definition 67
4.6.2 Request Validation and System Verification 68
4.7 Integrations 68
4.7.1 Intelligent Orthanc Plugin 69
4.7.2 Browser Extension 70
4.8 Chapter Summary 71
Chapter 5. Results 74
5.1 Contextual Inquiry Findings (RQ1) 74
5.1.1 Survey of Clinicians 75
5.1.1.1 Demographic Characteristics 75
5.1.1.2 Clinical Experience and Confidence Levels 79
5.1.1.3 Perceptions of Workload and Time Pressure 81
5.1.1.4 Challenges to Effectiveness and Efficiency 84
5.1.2 Interviews with Medical Doctors and Open Ended Survey Responses 86
5.1.2.1 Theme 1: Infrastructure-Driven Inefficiency: Suboptimal Access and Management of Digital Images 87
5.1.2.2 Theme 2: Workflow Fragmentation: Barriers to Effectiveness and Critical

Delays 88
5.1.2.3 Theme 3: Cognitive Overload: The Burden of Research and Consultative Support 89
5.1.2.4 Theme 4: Criteria for Intervention: Requirements and Limitations of Digital and Automated Tools 90
5.1.2.5 Themes from Open-Ended Survey Comments 91
5.1.2.6 Summary of Contextual Inquiry Findings 92
5.2 AI Model Performance Metrics (RQ2) 92
5.2.1 Classification Model Performance Results 93
5.2.2. Detection Model Performance Results 94
5.3 Experimental Evaluation Results (RQ3) 95
5.3.1 Participant Demographics 95
5.3.2 Results of workload assessment 97
5.3.2.1 Overall Workload Assessment 97
5.3.2.2 Subscale Analysis: Cognitive vs. Procedural Load 98
5.3.3 User Acceptance and Usability (TAM 2) 100
5.3.3.1 Construct Analysis 101
5.3.3.2 Distribution of Sentiment 102
5.3.4 Qualitative Feedback and Feature Requests 104
5.3 Chapter Summary 107
Chapter 6. Discussion 107
6.1 The Crisis of ‘Object’ and ‘Community’: Systemic Breakdowns in Radiological Workflows 109
6.1.1 The ‘Analog’ Gap: Information Flow as a Community Breakdown 109
6.1.2 Infrastructure as a Limiting Artifact: The Impact of Non-Dedicated Hardware 110
6.2 The AI Tool as a "Mediating Artifact": Resolving Workload Contradictions 111
6.2.1 Cognitive Offloading: Resolving the Subject-Object Tension 112
6.2.2 Scaffolding for Generalists: Bridging the Competence Gap 112
6.3 The Safety Philosophy of ‘Recall over Precision’: Technical Trade-offs in Context 113
6.3.1 The Clinical Rationale for High Recall (Sensitivity) 113
6.3.2 Defending Low Precision: AI as a ‘Triage Tool’ vs. ‘Gold Standard’ 114
6.4 Division of Labor and Professional Guardrails: The Ethics of Implementation 115
6.4.1 Shifting Roles: The Generalist as the New Primary Interpreter 115
6.4.2 The ‘User Competence’ Criterion: Preventing Deskilling 116
6.5 Limitations 117
6.5.1 Architectural Trade-offs and Model Justification 117
6.5.2 Training Data and Algorithmic Bias 118
6.5.3 Clinical Scope and Outcome Measurement 118
6.5.4 Experimental Design and Ecological Validity 118

Chapter 7. Conclusion 120
7.1 Synthesis of Findings: Resolving the Crisis of Capacity 120
7.2 Major Contributions and Significance 121
7.3 Recommendations for Policy and Practice 122
7.4 Recommendations for Future Work 123
7.5 Concluding Remarks 123
Chapter 8. References 124
Appendices 131
Appendix A: Semi Structured Interview Guide 131
Appendix C: NASA Task Load Index 132
Appendix D: In person Observation Plan 133
Appendix E: Perceived Usefulness and Ease of use Questionnaire 134
Appendix F: University of Zambia Biomedical Research Ethics Committee Approval 144
Appendix G: National Health Research Authority Ethical Clearance Approval 146
Appendix H: Ministry of Health Ethical Clearance Approval 147
Appendix I: University Teaching Hospitals Ethical Clearance Approval (1/2) 148
Appendix J: University Teaching Hospitals Ethical Clearance Approval (2/2) 149
Appendix K: Levy Mwanawasa Teaching Hospital Ethical Clearance Approval 150
Appendix L: Clinicians' Awareness and Perceptions of the Use of Artificial Intelligence in Medical Imaging in Zambia (Biodata and Challenges Sections) 151
Appendix M: Biodata and Consent form for Classification Tool Interaction 160
Appendix N: University of Zambia Biomedical Research Ethics Committee Approval Renewal 2023 165
Appendix O: University of Zambia Biomedical Research Ethics Committee Approval Renewal 2025 167

List of Tables

Table 2.1 Literature findings and gaps

Table 3.1 Chat elements and their application in the context of radiology in Zambia

Table 3.2 Development Timeline of the AI Interpretation Interface

Table 4.1 Detailed layer specification of the VGG16-based architecture

Table 4.2 Model Configuration and Hyperparameters

Table 4.3 Hyperparameter configuration for the RSNA Pneumonia Detection task

Table 4.4 Showing Key modifications introduced on recommendation in first focus group

Table 4.5 Showing Key modifications introduced on recommendation in second focus group

Table 5.1. Demographics of Clinicians Survey Participants

Table 5.2 Cross-tabulation of Experience with Complex Modalities (CT/MRI) versus Self-Reported Confidence Levels among Medical Practitioners.

Table 5.3 Pairwise Comparisons of Perceived Workload between Professional Groups (Dunn's Post-Hoc Test)

Table 5.4 Interview and Survey Open Response Participant Profile Table

Table 5.5 Performance Metrics for the Pneumonia Classification Model (VGG16).

Table 5.6 Performance Metrics for the Pneumonia Detection Model (Mask R-CNN)

Table 5.7. Demographics of Participants in the AI Tool Evaluation Experiment

Table 5.8. Paired Comparison of Global Workload Scores (Wilcoxon Signed-Rank Test)

Table 5.9 Kruskal Wallis test results for practice years and profession

Table 5.10 Summary of Technology Acceptance Model (TAM 2) Construct Scores

Table 5.11 Mann-Whitney U test of Gender comparison stats

Table 5.12 Kruskal-Wallis H test of perceived usefulness for practice years and Profession

Table 5.13 Experiment Open Response Participant Profile Table

List of Figures

Figure 1.1 Typical radiological workflow, with most stages done analogue in Zambia

Figure 3.1 CRISP-DM lifecycle as applied to the clinical AI research design, illustrating the six iterative phases from business understanding to deployment.

Figure 4.1 A high-level system diagram showing the layer interactions

Figure 4.2 The file selection module of the web interface supporting direct DICOM ingestion

Figure 4.3 Interface Prototype requiring manual triggers for inference

Figure 4.4 The v2.0 interface displaying automated detection results with standardized visualization

Figure 4.5. The final v3.0 interface features the multi-pathology suspicion panel

Figure 4.6 The Intelligent Orthanc Viewer plugin displays AI-detected pneumonia regions within the native DICOM viewer interface showing model results in the metadata panel

Figure 4.7 The intelligent browser extension analyses a chest radiograph within a Chrome browser session, featuring the metadata side-panel

Figure 5.1 Charts of Likert responses to workload and effectiveness statements

Figure 5.2 Radar chart illustrating the mean scores for the six NASA-TLX subscales

Figure 5.3 Diverging stacked bar chart illustrating the distribution of participant agreement levels across the core TAM 2 constructs. The dominance of the green bars (Agree/Strongly Agree) indicates widespread acceptance

List of Abbreviations in Alphabetical Order

Abbreviation	Description
AI	Artificial Intelligence
AP	Anterior to Posterior
API	Application Programming Interface
AUPRC	Area Under the Precision-Recall
CHAT	Cultural-Historical Activity Theory
CRISP-DM	CRoss Industry Standard Process for Data Mining
CT	Computerised Tomography
CXR	Chest X-Ray
DICOM	Digital Imaging and Communications in Medicine
DOM	Document Object Model
EHRs	Electronic Health Records
EI	Enterprise Imaging
EMI	Enterprise Medical Imaging
EMIZ	Enterprise Medical Imaging In Zambia
FDA	United States Food and Drug Administration
FPN	Feature Pyramid Network
GRZ	Government of the Republic of Zambia
HCI	Human-Computer Interaction
IoU	Intersection over Union
LMIC	Low and Middle Income Country
LMUTH	Levy Mwanawasa University Teaching Hospital
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NGO	Non Governmental Organization

PA	Posterior to Anterior
PACS	Picture Archiving and Communication System
RIS	Radiology Information System
RQ	Research Question
RSNA	Radiological Society of North America
TAM	Technology Acceptance Model
UNZA	The University of Zambia
UTHs	University Teaching Hospitals
ViT	Vision Transformer

Chapter 1. Introduction

1.1 Opening Context and Background

Radiological analysis constitutes a fundamental component of modern medical diagnoses, essential for interpreting a wide range of conditions from trauma-related fractures to antenatal ultrasounds. However, there exists a shortage of radiological expertise around the world. Further compounding this is the fact that the global distribution of radiological expertise is highly uneven. This results in critical shortages across many developing regions, particularly the Global South [118]. In Zambia, this challenge is particularly acute; reports from 2020 indicated that only eight trained radiologists were operational within the public sector, translating to an estimated density of 0.5 radiologists per 1 million people [63]. This scarcity severely strains the capacity of the public healthcare system to provide timely and accurate diagnostic imaging services. The current workload is disproportionately carried by a small cohort of overburdened specialists and non-radiologist medical practitioners who must frequently interpret complex images without adequate support. Efforts to mitigate this challenge typically involve strategies such as recruiting and training additional radiologists as well as investment in equipment and other tooling.



Figure 1.1 Typical radiological workflow, with most stages done analogue in Zambia.

Source: Generated with AI.

In the absence of specialists, the burden of interpretation shifts to the Generalist Medical Doctor. These practitioners, often stationed in rural or high-volume urban facilities, are forced to step into a specialized role for which they have received limited training. The operational reality they face is characterized by friction and fragmentation. As illustrated in Figure 1.1, the typical workflow for Zambian public facilities, the process is often analogue and disjointed: a patient physically carries a handwritten request form and a CD containing their X-ray from the imaging department to the doctor's consultation room. The doctor must then attempt to load these images on non-dedicated hardware, often without access to the patient's full clinical history. This creates a high-pressure environment where a non-specialist is expected to make life-altering decisions based on complex visual data, often in isolation. It is within this context of an overburdened workforce and a fragmented workflow that the need for innovation becomes urgent. The central premise of this research is not simply to automate diagnosis, but to provide a supportive scaffold for these generalist practitioners.

This research focuses on the fourth stage in the workflow (image and data interpretation).

This research posits that an immediate, impactful strategy lies in enhancing the productivity of existing medical practitioners through the integration of advanced technological tools, specifically, Artificial Intelligence (AI) solutions. Leveraging recent exponential growth in computing power and data availability presents a viable opportunity to augment human capability in medical image interpretation. This study was conducted as a critical component of the larger Enterprise Medical Imaging in Zambia (EMIZ)[1] , an ongoing initiative exploring the deployment of EMI and AI to streamline effective radiological workflows in Zambian public health facilities [24].

1.2 Problem Statement

The core problem facing the Zambian radiological sector is a systemic imbalance between the complexity of diagnostic tasks and the capacity of the available workforce. While the volume of medical imaging is increasing, the scarcity of radiologists has created a bottleneck where

[1] https://emi.org.zm/

complex interpretation tasks are routinely offloaded to generalist doctors. This shortage of experts causes longer turnaround times and worse patient outcomes.

1.3 Research Gap and Justification

While the transformative potential of AI in medical imaging demonstrated through successful classification, segmentation, and localization models globally is slowly becoming mainstream, there remains a substantial gap in the practical application and evaluation of these tools within the specific operational and resource constraints of Zambian radiological workflows. Much of the AI research utilizes large, publicly available datasets predominantly sourced from Western populations, raising critical concerns regarding model generalizability and algorithmic bias when deployed in an African healthcare context [106]. This research is justified by the need to move toward validated clinical utility. Specifically, there is a scarcity of studies that evaluate the usability, usefulness, and long-term adoption of AI tools tailored to the unique challenges faced by practitioners in Sub Saharan Africa, Zambian practitioners in particular and challenges in implementing these AI solutions. These challenges include the pervasive use of non-dedicated viewing devices (e.g., mobile phones), difficulties in acquiring annotated local data and severe overburdening workload among others [72, 106]. This study addresses this gap by moving beyond model performance metrics to evaluate systemic integration. By employing the CHAT [107] framework, this research analyses the integration of AI not as a plug-and-play solution, but as a "system-altering artifact." CHAT provides the necessary lens to identify and resolve the contradictions between the Subject (the generalist) and the Community (the fragmented hospital system). This justification grounds the study in the need to move from "theoretical algorithm performance" to "verified clinical utility" in a resource-constrained environment.

1.4 Research Aim

The overarching aim of this research is to design, implement, and evaluate the integration of AI assistive tools within the radiological workflows of Zambian public health facilities, with the goal of mitigating practitioner workload and improving diagnostic confidence in a resource-constrained environment.

1.5 Research Objectives and Questions

This research was structured around three interconnected objectives, which guided the methodological design and subsequent analysis:

1.5.1 Research Objectives

1. To understand the existing challenges in medical image interpretation workflows relative to workload, efficiency, and effectiveness in the Zambian context.

2. To explore and create practical AI solutions that support medical practitioners, taking into account relevant imaging modalities and prevalent local pathologies.

3. To investigate how to rigorously evaluate the devised AI solutions based on biases and workflows of the different cadres of practitioners.

1.5.2 Research Questions

The following research questions (RQ) were formulated to guide the inquiry:

1. What challenges exist in radiological image interpretation relative to workload, effectiveness, and efficiency within Zambian public health facilities?

2. What AI solutions can be effective and adaptable for mitigating the challenges faced by medical practitioners in the Zambian context?

3. What specific criteria should be used to select the final AI tools for reliable and sustainable implementation in Zambian radiological workflows?

1.6 Significance of the Study

This study holds substantial significance for Zambia's healthcare system and for broader AI adoption across LMICs. By systematically analysing the operational, infrastructural, and workflow challenges present in Zambian radiology, the research provides contextually grounded evidence required to guide responsible implementation of AI-assisted diagnostic tools. Furthermore, by designing and evaluating AI solutions tailored to local modalities, data constraints, and practitioner needs, this work contributes practical insights for developing assistive technologies that are both usable and scalable within resource-limited environments. The proposed framework for AI integration, including criteria such as transparency, contextual relevance, workload alignment, and compatibility with existing systems offers actionable

guidance for policymakers, health administrators, and technologists seeking to deploy AI tools at national or institutional levels.

1.7 Thesis Outline

This manuscript is organized into six chapters as follows:

- Chapter 1 is the introduction, outlining background, research objectives, research questions, and the scope of the study.
- Chapter 2 presents a comprehensive review of the related literature, focusing on the global challenges in medical image interpretation, the history and current landscape of AI interventions in radiology, clinical validation methods, and the regulatory and ethical considerations pertinent to the Global South.
- Chapter 3 details the research methodology, including the application of the CHAT [83] framework, the convergent mixed-methods approach, the implementation of the AI classification and detection models for pneumonia, and the design of the controlled evaluation experiments.
- Chapter 4 describes the system design and implementation, documenting the technical architecture of the AI models and the development of the user-centric web interface.
- Chapter 5 reports the findings from the clinician surveys, thematic analysis of the interviews, technical performance of the AI models, and the results from the workload and usability evaluation.
- Chapter 6 discusses these findings, interpreting the results within the CHAT framework, addressing the research questions, discussing the study's limitations, and comparing the outcomes to existing literature.
- Chapter 7 concludes the thesis, summarizing the major contributions, and offering recommendations for future work and long-term AI adoption strategies in Zambia.

1.8 Chapter Summary

This chapter provided the foundational context for the study, highlighting the critical shortage of radiological expertise in Zambia and the resulting operational strain placed on generalist

medical practitioners. It defined the core problem of diagnostic bottlenecks and identified a substantial research gap regarding the practical, validated application of AI tools in Sub-Saharan African clinical workflows. Furthermore, the chapter outlined the overarching aim, specific research objectives, and the significance of the study in transitioning AI from theoretical performance to clinical utility. The next chapter will explore the existing literature surrounding global radiological challenges, AI interventions, and the theoretical frameworks guiding this research.

Chapter 2. Related Work

This chapter presents a critical review of the literature underpinning the development and evaluation of AI interventions in radiological workflows. Rather than providing a general history of medical computing, this review focuses on the convergence of three distinct domains: the clinical constraints of radiology in low-resource settings, the state-of-the-art in deep learning architectures for medical image analysis, and the sociotechnical frameworks used to evaluate human-AI interaction.

The review is structured to build the theoretical and technical justification for the study's methodology. Section 2.1 examines the specific operational challenges in the Global South that necessitate technological intervention. Section 2.2 critiques current deep learning approaches, specifically motivating the selection of Transfer Learning and Object Detection architectures for resource-constrained environments. Section 2.3 analyses evaluation metrics and frameworks, arguing for a shift from pure algorithmic accuracy to clinical utility measures such as cognitive workload and user acceptance. Finally, Section 2.4 addresses the ethical and socio-technical dimensions of deploying AI in environments characterized by data scarcity and infrastructural deficits.

2.1 The Clinical Context: Radiology in Resource-Limited Settings

Radiological image interpretation faces a number of challenges around different areas. Our work will contribute to this by centering around Zambia, and around efficiency and workload. We endeavour to discuss the challenges faced here.

2.1.1 Shortage of radiologists

Radiological image interpretation faces a critical shortage of skilled experts, both globally and particularly in regions like the Global South and Zambia. Studies [48, 74] have highlighted the scarcity of radiologists, posing significant challenges to timely and accurate diagnosis. In the context of Zambia, this shortage is made worse by various factors, including brain drain of skilled professionals [61]. This lack of radiologists not only impacts the efficiency of healthcare delivery but also contributes to increased patient waiting times and diagnostic delays, ultimately compromising patient outcomes.

Across the globe, there exists a persistent shortage of radiologists, with demand far outstripping supply. According to the Radiological Society of North America (RSNA)[2] Europe had 13 radiologists per 100,000 people with the U.K. having 8.5 per 100,000. Factors contributing to this shortage include an ageing workforce [44], growing demand for imaging services [44], and limited exposure and interest from doctors in training to become radiologists. This imbalance between supply and demand underscores the urgent need for innovative solutions to bridge the gap and ensure access to timely and accurate diagnostic services worldwide. Kalidindi and Gandhi [44] propose AI as a potential solution but don't go into details on how such solutions would work and look.

Regions in the Global South, including sub-Saharan Africa, face particularly acute shortages of radiologists [40, 74]. The disparity in healthcare infrastructure and investment exacerbates this challenge, limiting access to essential diagnostic services for millions of people. According to Tahir et al [97] Sub Saharan Africa " … Averages fewer than 22 doctors per 100,000 people." They further mention that there's about 1 radiologist per 600,000 people in Nigeria for example. They go on to suggest telemedicine as a solution to this issue which could be effective.

In Zambia, the shortage of radiologists is a pressing concern , with a limited number of trained professionals available to meet the growing demand for imaging services [12]. An article in the Pan African Medical Journal in 2020 showed that there were 8 trained radiologists operating in the public sector in Zambia [63] with 0.5 radiologists to 1 million people [63].

[2] https://www.rsna.org/news/2022/may/global-radiologist-shortage

This scarcity is compounded by factors such as rural-urban migration, inadequate infrastructure, and resource constraints.

2.1.2 Infrastructure and resource constraints

Infrastructure and resource constraints add significant barriers to patients' ability to access quality radiological services. Limited funding for healthcare infrastructure, outdated equipment and inadequate support systems [13, 116] hinder the delivery of imaging services. With the Zambia National Health Strategic Plan 2017 - 2021 stating that 60 percent of equipment in 2016 was old and outdated. They also impede the recruitment and training of skilled professionals. Addressing these constraints will require strategic investments in healthcare infrastructure, technology upgrades and capacity building initiatives to create an enabling environment for radiological practice and ensure high quality care to all Zambians.

2.1.3 Challenges with Imaging data

Imaging data and its quality presents a set of challenges in radiological workflows, ranging from equipment limitations to data storage constraints [3]. These challenges can impact the quality and availability of diagnostic images, potentially compromising patient care and clinical decision-making.

Equipment limitations, including outdated technology, maintenance issues, and limited access to advanced imaging modalities, pose significant challenges to acquiring high-quality diagnostic images in Zambia [116]. Inadequate equipment can result in suboptimal image quality, diagnostic errors, and delays in patient care. Addressing equipment limitations requires investment in modern imaging technology, regular maintenance and calibration, and training programs to ensure optimal performance and reliability of radiological equipment across healthcare facilities in Zambia.

This workforce crisis is exacerbated by a severe infrastructure deficit. Unlike high-resource settings where Picture Archiving and Communication Systems (PACS) are ubiquitous, radiological workflows in the Global South often rely on ‘analogue’ data transfer. As noted by Mollura et al. [65], the absence of integrated RIS/PACS means that patient data is frequently siloed on physical media (CDs) or hard films.

A critical, often overlooked challenge in this context is the phenomenon of 'secondary capture.' Due to the lack of diagnostic-grade monitors, clinicians frequently photograph X-ray films using mobile phones or view them on non-calibrated screens in variable lighting conditions. This practice fundamentally degrades the image quality, introducing artifacts such as glare, geometric distortion, and reduced contrast resolution. Consequently, any AI intervention designed for this context cannot assume 'perfect' DICOM inputs; it must be robust enough to handle the noise and variability inherent in this survival workflow.

The combination of workforce scarcity and infrastructural friction creates a dangerous bottleneck. Generalist doctors, lacking specialized training, report high levels of diagnostic uncertainty when interpreting complex modalities like CT or subtle chest radiographs. The literature suggests that in such settings, the primary need is not for "autonomous diagnosis" (which raises ethical and legal barriers) but for decision support systems that can triage cases, highlight abnormalities, and provide a "second opinion" to reduce cognitive load [3]. This distinction defines the functional requirements for the AI tool proposed in this study: it must function as a safety net for the generalist, rather than a replacement for the specialist.

Data and storage limitations present additional challenges in managing and accessing imaging data in radiological workflows. Limited storage capacity, outdated archiving systems, and interoperability issues can impede the efficient storage, retrieval, and sharing of medical images, hindering collaborative decision-making and continuity of care [118]. To address these limitations, healthcare institutions must invest in robust data storage infrastructure, implement standardised protocols for image acquisition and management, and adopt interoperable imaging systems to facilitate seamless data exchange and integration within the healthcare ecosystem.

2.2 Comparative Studies and Lessons from Similar Contexts

2.2.1 AI in Medical Imaging in Other LMICs

Pilot programs in LMICs have demonstrated the potential of AI in improving healthcare outcomes. For instance, the implementation of AI-enabled foetal monitoring software at the Area 25 Health Centre in Lilongwe led to a reduction in stillbirths and neonatal deaths over three years [14]. Other AI interventions in LMICs included, non-sonographer healthcare workers could be trained in AI-based ultrasound triage, potentially reducing the workload of

formally trained sonographers. COVID-19 triage became more time-efficient with an AI CT-screening tool [16]. Such successes underscore the feasibility of AI interventions in similar resource-constrained environments.

2.2.2 Successes and Failures

Conversely, the deployment of AI technologies without adequate consideration of local contexts has led to failures. Challenges such as inadequate infrastructure, lack of trained personnel, regulation around data and cultural resistance have impeded the successful adoption of AI in some LMICs [3, 49]. These experiences highlight the importance of tailoring AI solutions to the specific needs and capacities of the target healthcare systems.

Table 2.1 Literature findings and gaps

Author(s)	Finding	Gaps
H. Hah and D. S. Goldin	Although sentiment was positive, the current form of AI assistance may not be linked with efficient decision making, as AI algorithms are not well aligned with subjective human reasoning in clinical diagnosis.	This study was conducted on clinicians from the Florida International University, and the perceptions and challenges are based on that resource setting.
J. Stewart *et al.*	Emergency clinicians interviewed were generally optimistic about the use of AI in emergency medicine, so long as it is used as a decision support tool and they maintain the ability to override its recommendations.	These perceptions are from participants in the high resource Australian Healthcare service.
A. Karera, F. Davidson, and P. Engel-Hills	Study suggests the formal introduction of radiographers into image interpretation.	While part of this study was conducted in Namibia, while this is a lower resource environment than the west it is a different context from Zambia.
D. J. Mollura *et al*	This study highlights the significant radiology gaps globally, and suggests AI has	This study doesn't endeavour to implement these systems for the

	unique potential in LMICs and further highlights challenges to AI adoption in these settings	practitioners outside training environments.

2.3 Technical State-of-the-Art: Deep Learning in Medical Imaging

The application of Artificial Intelligence in radiology has shifted from experimental feature engineering to the deployment of robust Deep Learning architectures. This section reviews the specific architectures relevant to this study, Convolutional Neural Networks (CNNs) and Region-Based CNNs and justifies their selection for resource-constrained medical environments.

Before delving into specific architectures, it is crucial to recognize the broader landscape of medical AI. Recent comprehensive surveys have extensively documented the evolution of AI in medical imaging, highlighting a rapid transition from basic machine learning to deep neural networks [55]. In the specific domain of radiology, systematic reviews demonstrate that AI applications are predominantly focused on chest radiography and neuroimaging, with performance often matching expert human readers in controlled settings [39]. Furthermore, literature focusing on AI for chest radiograph interpretation and pneumonia detection emphasizes that deep learning has become the standard for addressing diagnostic bottlenecks [45]. Finally, scoping reviews on transfer learning in medical imaging confirm that leveraging models pre-trained on large datasets is the most robust method for overcoming the limited dataset sizes frequently encountered in clinical research [67]. These broad surveys establish the foundational efficacy of CNNs, providing the justification for selecting VGG16 [93] and Mask R-CNN for this study.

2.3.1 Deep Learning Classification

The application of CNNs to medical image analysis is often constrained by the 'data scarcity' problem. Unlike general computer vision tasks, where datasets like ImageNet [18] contain millions of labelled examples, medical datasets (particularly in low-resource settings) are often small, fragmented, and expensive to annotate. To overcome this limitation, the literature overwhelmingly supports the use of Transfer Learning.

Transfer learning leverages the feature representations learned by a model on a large source domain (e.g., natural images) to solve a task in a target domain (e.g., medical X-rays) with significantly less training data. Research by Tajbakhsh et al. [98] demonstrated that fine-tuning pre-trained deep CNNs consistently outperforms training from scratch, particularly when the target dataset is limited. In the context of pneumonia detection, this approach allows the model to repurpose low-level feature detectors (identifying edges, curves, and textures) learned from natural images to identify radiological features like lung opacities and consolidations

Among the various CNN architectures, VGG16 [93], proposed by Simonyan and Zisserman, remains a cornerstone for transfer learning in radiology. Characterized by its uniform architecture of 13 convolutional layers and 3 fully connected layers, VGG16 utilizes small 3 by 3 convolution filters to capture increasingly complex spatial hierarchies.

While newer architectures like Efficient Net or Vision Transformers offer higher theoretical efficiency, VGG16 is frequently selected in medical AI research. Numerous comparative studies in chest radiography utilize VGG16 as a reliable benchmark due to its proven feature extraction capabilities and transfer learning efficiency. Its deep architecture allows for the robust extraction of hierarchical features, which is critical for distinguishing subtle pathological textures in grayscale X-rays. In these studies, VGG16 frequently achieves sensitivity levels comparable to human radiologists [80], establishing it as a standard and highly stable baseline against which newer, more complex models are continuously evaluated.

Disease classification is also a very important use case for AI models and approaches. Classification ‘is this lesion normal/abnormal, or benign/malignant?’ is a major part of the diagnostic process. AI classification is one of the more widely used techniques with it commonly being used with x rays and other modalities.

2.3.2 Object Detection and Segmentation: Review Mask R-CNN

There have been a number of systems built to assist in the diagnoses of various pathologies using AI [30] and some built to fully automate it [56, 59, 78, 115]. These have shown the ability for AI to meet and in some cases exceed the performance of humans in these applications [54, 57, 60]. Therefore AI presents an option to reduce the severe challenge the shortage of radiologists presents.

Image segmentation which is ‘delineation of an organ/tissue/lesion of interest’ [94] is a critical application of AI in medical imaging, focusing on the partitioning of medical images into meaningful regions or segments . This process facilitates more detailed analysis and quantification of anatomical structures and pathological findings, enhancing the diagnostic and treatment planning capabilities of radiologists.

AI algorithms, particularly those based on deep learning, have shown remarkable success in image segmentation tasks. Convolutional Neural Networks (CNNs) [112], especially U-Net architectures [101], are widely used due to their ability to capture both high-level features and spatial details in medical images [27, 35, 55]. These networks are trained on annotated datasets, learning to identify and delineate structures such as organs, tumours, and blood vessels with high precision.

There are many uses of image segmentation included but not limited to tumour segmentation, organ segmentation and vascular segmentation.

While binary classification (Normal vs. Pneumonia) provides a useful triage signal, clinical intervention requires precise **localization**. A simple "Positive" label without spatial context can be a "black box" prediction, failing to engender trust in the clinician. To address this, current research emphasizes the transition from classification to **Object Detection** and **Instance Segmentation**.

The state-of-the-art architecture for this task is **Mask R-CNN** [34]. Developed as an extension of Faster R-CNN, this framework introduces a critical innovation: a parallel branch for predicting segmentation masks in addition to the existing branches for bounding box coordinates and class labels.

Mask R-CNN operates in two stages:

1. **Region Proposal Network (RPN):** It scans the feature map to propose "Regions of Interest" (RoIs) that likely contain an object.
2. **RoIAlign and Head Prediction:** It extracts features from these regions using a precise alignment mechanism (RoIAlign) to generate pixel-accurate masks and bounding boxes.

In the context of chest radiography, pathologies like pneumonia often manifest as diffuse, amorphous opacities rather than distinct objects with rigid boundaries. Mask R-CNN is particularly well-suited for this challenge because it combines object detection (locating the general area) with segmentation (delineating the specific shape).

Using Mask R-CNN transforms the AI tool from a simple classifier into an **'Augmented Intelligence'** assistant. By visualizing the suspicious region with a bounding box, the model provides an inherent form of explainability. This supports the "Human-in-the-Loop" workflow by directing the clinician's attention to the specific area of concern, allowing them to validate or reject the AI's findings based on visual evidence. This capability is essential for the "cognitive offloading" function required in high-volume settings, as it reduces the visual search time for the radiologist.

2.3.3 AI in the Global South: Challenges of Deployment

The deployment of Deep Learning models in the Global South faces unique challenges often overlooked in mainstream research. A primary concern is Domain Shift [28], the degradation in model performance when the training data (Source Domain) differs statistically from the real-world deployment data (Target Domain).

Most publicly available datasets, such as MIMIC-CXR [43] or ChestX-ray14 [110], are sourced from well-resourced Western hospitals using state-of-the-art digital detectors. In contrast, radiographs in Zambian facilities are often captured using older analogue machines or are digitized via "secondary capture" (photography). This discrepancy creates a significant data drift, where models trained on 'clean' Western data may fail to generalize to the 'noisy' local images [106].

Furthermore, the Data Scarcity in the region (specifically the lack of annotated local datasets) hinders the development of region-specific models. The literature suggests that to mitigate these risks, AI interventions in Africa must prioritize robustness over raw accuracy [50]. Strategies such as heavy data augmentation (simulating noise and rotation) and the use of "Human-in-the-Loop" validation protocols are essential to ensure patient safety when transferring models across these distinct domains [79, 117].

Since the release of ChatGPT[3] there has been a new flush of AI chatbots with various use cases. There is potential for some use cases to be found in radiology and other aspects of healthcare [2, 8, 17]. Researchers have written about the potential and issues of such applications.

2.4 Human-AI Interaction and Evaluation Frameworks

Evaluating Artificial Intelligence in healthcare requires moving beyond algorithmic performance metrics to assess the system's impact on the human practitioner. A technically accurate model may fail clinically if it increases workload, disrupts established workflows, or lacks user trust. Therefore, a comprehensive evaluation must triangulate three dimensions: Clinical Performance Metrics (to validate diagnostic safety), Workload Assessment (to measure cognitive cost), and User Acceptance (to predict adoption).

2.4.1 Metrics for Clinical AI: The Accuracy Paradox

In medical machine learning, standard metrics such as Accuracy can be dangerously misleading, particularly in unbalanced datasets where the prevalence of disease is low [36, 95]. For example, in a dataset where 95 percent of patients are healthy, a model that simply predicts 'Healthy' for every case would achieve 95 percent accuracy while missing every single pathology. This phenomenon, known as the 'Accuracy Paradox,' necessitates a shift toward metrics that capture the clinical consequences of error. To understand these trade-offs, it is essential to define the foundational metrics based on True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN):

Accuracy: The proportion of total predictions that were correct.

$$Accuracy = \frac{TP + TN}{Total\ Number\ of\ Instances}$$

Precision (Positive Predictive Value): The proportion of positive predictions that are pathological.

$$Precision = \frac{TP}{TP + FP}$$

[3] https://openai.com/index/chatgpt/

Recall (Sensitivity): The proportion of actual positive cases that were correctly identified by the model.

$$Recall = \frac{TP}{TP + FP}$$

F1-Score: The harmonic mean of Precision and Recall, which provides a balanced measure when classes are uneven.

$$F1 - Score = \frac{Precision \times Recall}{Precision + Recall}$$

Recent literature on AI in resource-limited settings emphasizes the prioritization of Recall over Precision. In a triage or screening workflow, such as identifying pneumonia in a rural clinic without a radiologist, the primary objective is to minimize False Negatives. As argued by identifying Oakden-Rayner et al. [76], a missed diagnosis in a setting with limited follow-up capacity can be catastrophic.

Conversely, a lower precision (leading to more False Positives) is often considered an acceptable trade-off in "Human-in-the-Loop" systems. When an AI tool acts as a "safety net" rather than an autonomous diagnostician, a false positive merely prompts a human review, whereas a false negative may result in a patient being discharged without treatment. Therefore, effective evaluation of clinical AI must focus on the F1-Score and the Area Under the Precision-Recall Curve (AUPRC), rather than simple accuracy, to ensure the model is optimized for patient safety [88].

2.4.2 Measuring Cognitive Load and Acceptance

To evaluate the human impact of the AI intervention, this study employs two established psychometric instruments: the NASA-TLX and the TAM 2. The selection of these frameworks is grounded in their proven efficacy within medical informatics and Human-Computer Interaction (HCI) research.

The NASA-TLX, originally developed by Hart and Staveland [31], is the gold standard for assessing subjective workload across six dimensions: Mental, Physical, and Temporal

Demands, Performance, Effort, and Frustration. While alternatives like the Subjective Workload Assessment Technique exist, NASA-TLX has demonstrated superior sensitivity in distinguishing workload variations in high-stakes healthcare tasks [86].

In the context of radiology, where 'decision fatigue' is a critical bottleneck, NASA-TLX allows researchers to pinpoint the specific type of load an AI tool affects. For instance, studies by Grier [29] and others in medical HCI have used NASA-TLX to demonstrate that while AI may reduce *time* pressure (Temporal Demand), it can inadvertently increase *cognitive* pressure (Mental Demand) if the interface is poorly designed. This multidimensional granularity is essential for answering Research Question 2 regarding the tool's impact on efficiency versus effort.

To predict clinical adoption, the **TAM 2** [105] was selected over alternatives like the **Unified Theory of Acceptance and Use of Technology (UTAUT)** [104]. TAM 2 extends the original model by incorporating social influence processes (Subjective Norm) and cognitive instrumental processes (Job Relevance), which are critical in hierarchical medical environments.

Previous studies in digital health adoption in the Global South [4] have successfully employed TAM 2 to identify that **Perceived Usefulness (PU)**—the degree to which a tool enhances job performance—is often the strongest predictor of intention to use among physicians. By using TAM 2, this study aligns with a robust body of literature in healthcare informatics that links system characteristics (like output quality) directly to professional acceptance [105], allowing for comparability with existing telemedicine and e-health evaluations [46].

2.5 Regulatory and Ethical Considerations

2.5.1 Data Privacy and Patient Consent in AI Diagnostics

The integration of AI in radiology necessitates handling sensitive health data, often involving identifiable medical images. Globally, privacy regulations such as the General Data Protection Regulation (GDPR) [38] and Health Insurance Portability and Accountability Act (HIPAA) [22] emphasize informed consent, anonymization, and data minimization, but similar frameworks are either absent or underdeveloped in the global south. In Zambia, as in many

African nations, comprehensive data protection laws are still evolving. Patient consent protocols are often informal, and national health data protection frameworks like Zambia's Data Protection Act of 2021[4] remain under-implemented. This poses risks regarding AI systems trained on patient data without stringent anonymization or explicit consent, highlighting the urgent need for localized ethical guidelines. Research has been done on the implications of AI on healthcare [66, 71, 111] looking into the challenges protecting the data and potential solutions.

2.5.2 Ethical Concerns Around AI Decision-Making in Healthcare

AI decision systems raise ethical questions, especially in contexts where oversight may be limited. For instance, the risk of algorithmic bias stemming from non-representative training datasets can disproportionately affect African populations. Work has shown that biases in AI training and tools can negatively affect minorities and underrepresented groups [21]. The ethical and fairness issues have been discussed by many [5, 68–70, 75, 103].

2.5.3 Regulations Around AI in Healthcare

Globally, regulatory bodies like the U.S. FDA[7, 102] and EU MDR [25] are developing pathways for approving AI-based medical tools. However, in Zambia and similar countries, there is a vacuum in regulatory readiness. AI solutions are typically implemented through partnerships or pilot programs without formal national regulation [6, 15]. This lack of oversight could lead to unvalidated tools influencing clinical decisions, underscoring the importance of building health-tech regulatory capacity in the country. It is important to note however, that even these regulators are running projects and research to expand their coverage and regulation of AI in Healthcare [1, 26, 64, 77, 81].

2.6 Datasets and Benchmarking in Medical Imaging AI

2.6.1 Publicly Available Datasets

Many of the most commonly used datasets in AI imaging research like ChestX-ray 14 [109] and MIMIC-CXR [43], have driven significant advancements in AI imaging, they are heavily skewed toward Western demographics and high-resource clinical environments. Recognizing

[4] https://www.parliament.gov.zm/node/8853

this limitation, there has been a recent push to release non-Western datasets. Notable examples include VinDr-CXR from Vietnam [73], paediatric-specific datasets for pneumonia [52], the TBX11k dataset for tuberculosis [58], and the JSRT database from Japan [92].

Although these datasets improve global representation, they do not entirely bridge the gap for Sub-Saharan Africa. The critical issue lies in the differences in disease prevalence and epidemiological presentation. For example, while Asian datasets may feature high incidences of specific tuberculosis strains, the co-morbidity profile in Zambia, often complicated by high rates of HIV/AIDS, endemic parasitic infections, and specific bacterial pneumonia strains presents a fundamentally different radiological landscape. Furthermore, these non-Western datasets are still often captured using modern digital detectors, failing to represent the localized "survival workflows" in Zambia where older, analogue equipment and "secondary capture" (photographing films) are prevalent. Consequently, models trained on these datasets still risk severe domain shift when deployed in Zambian facilities.

2.6.2 Limitations of Existing Datasets for Global South Representation

The scarcity of high-quality, annotated datasets from Zambia or the broader Global South hinders the development of locally-relevant models [3, 114]. In the public sector storage of images tends to be in analogue means making it difficult to train data off these images [118]. Furthermore, low bandwidth and limited storage infrastructure make it difficult to collect and curate datasets for training. To address this, approaches like federated learning, synthetic data generation, and public-private partnerships are being explored globally but remain underutilized locally.

2.6.3 Challenges in Acquiring Annotated Data

In Zambia, the scarcity of trained radiologists poses a significant hurdle in generating annotated medical imaging datasets. The annotation process is time-consuming and requires expert knowledge, making it challenging to develop localized datasets necessary for training AI models suited to the Zambian healthcare context.

2.7. Human-AI Collaboration

2.7.1 Radiologist-in-the-Loop Systems

Integrating AI as an assistive tool rather than a replacement for radiologists can enhance diagnostic accuracy and efficiency. Radiologist-in-the-loop systems allow for human oversight, ensuring that AI-generated interpretations are reviewed and validated by medical professionals, thereby maintaining clinical accountability. Startups like Mecha health[5] are advancing radiologist in the loop applications. Their team is working on next generation foundation models for radiology, including topics such as producing complete reports and direct scan analysis.

2.7.2 AI as a Triage Tool vs. Decision Support

In resource-limited settings like Zambia, AI can serve as a triage tool to prioritize cases requiring immediate attention, thereby optimizing workflow and resource allocation. Alternatively, AI can function as a decision support system, providing second opinions to aid clinicians in diagnosis, especially in facilities lacking specialized radiologists.

2.8 Sustainability and Scalability of AI Solutions

2.8.1 Cost and Maintenance of AI Systems in Resource-Limited Settings

The implementation of AI systems entails significant costs related to procurement, maintenance, and training. In Zambia, where healthcare budgets are constrained, sustaining AI technologies post-deployment poses a considerable challenge. Strategies such as leveraging open-source AI tools and establishing partnerships with international organizations may alleviate some financial burdens.

2.8.2 Training and Local Capacity Building

For AI to be effectively integrated into Zambian radiological workflows, there is a need for comprehensive training programs targeting healthcare professionals. Building local capacity

[5] https://www.mecha-health.ai/

ensures that the workforce is equipped to operate, interpret, and maintain AI systems, thereby fostering self-reliance and reducing dependency on external expertise.

2.8.3 Long-Term Adoption and Support Challenges

Sustainable adoption of AI in Zambia requires alignment with national health strategies and policies. Integrating AI solutions into existing health information systems and ensuring regulatory compliance are critical for long-term success. Additionally, continuous monitoring and evaluation mechanisms should be established to assess the impact and efficacy of AI interventions over time.

2.9 Chapter Summary

This chapter provided a critical review of the literature surrounding the integration of Artificial Intelligence in radiological workflows, specifically contextualized for resource-constrained environments like Zambia. It highlighted the acute clinical challenges such as workforce shortages, analogue data transfer, and infrastructural deficits that necessitate AI intervention. A review of comparative studies from other LMICs underscored the importance of designing context-aware solutions rather than directly importing tools built for high-resource settings. Technically, the chapter justified the selection of robust deep learning architectures, specifically VGG16 for transfer learning and Mask R-CNN for object detection, as ideal candidates for addressing data scarcity and providing explainable diagnostic support.

Furthermore, the review identified the limitations of current, predominantly Western-centric datasets, highlighting the severe risk of domain shift when applying these models to local Zambian workflows. To evaluate these technical solutions effectively, this chapter established the need to move beyond raw algorithmic accuracy, advocating for psychometric and usability frameworks such as NASA-TLX and TAM 2 to measure cognitive workload and clinical acceptance. Finally, the regulatory, ethical, and sustainability challenges inherent in deploying AI in the Global South were discussed. Together, these elements form the theoretical and technical foundation for this study. Building on these insights, the next chapter will detail the research methodology, outlining the specific frameworks, data collection methods, and experimental design used to implement and evaluate the proposed AI solutions.

Chapter 3. Research Methodology

This chapter details the systematic research process employed to address the study's objectives. It outlines the methodological framework used to investigate the challenges in radiological workflows in Zambia, the strategy for developing Artificial Intelligence (AI) interventions, and the rigorous procedures used to evaluate these solutions.

The chapter is structured by adapting the "Research Onion" framework [89] to define the foundational research philosophy and design. For the operational execution of the study, specifically the development and deployment of the AI solution—the Cross-Industry Standard Process for Data Mining (CRISP-DM) lifecycle was adopted. This structure ensures a logical flow from the theoretical underpinnings (philosophy and approach) to the practical execution (data collection, modelling, and evaluation).

The chapter begins by defining the **Research Philosophy and Approach (3.1)** and the **Theoretical Framework (3.2)**. It then details the **Research Design (3.3)** before proceeding through the phases of the CRISP-DM lifecycle: **Contextual Inquiry (3.4)**, **Data Preparation (3.5)**, **Modelling Strategy (3.6)**, **Iterative Design (3.7)**, and **Evaluation Methodology (3.8)**. Finally, it addresses the **Ethical Considerations (3.9)** inherent in handling sensitive medical data.

3.1 Research Philosophy

Following the outer layers of Saunders et al.'s research onion, this section defines the underlying ontological and epistemological assumptions guiding the research and the methodological choice used to gather and analyze data.

3.1.1 Philosophical Stance

This study adopts a **pragmatic philosophical stance**. Pragmatism is particularly appropriate that seeks both context-specific understanding and measurable, practical outcomes. It posits that reality is complex and cannot be fully captured by a single philosophical position. Therefore, the most appropriate methods are those that best answer the research questions and address the practical problem at hand.

Within this pragmatic framework the research integrates two distinct perspectives:

1. **Positivism:** The **quantitative components of the study are informed by positivist assumptions**, namely that phenomena such as workload, usability, perceived usefulness and algorithm accuracy can be objectively measured through empirical observation. Accordingly, the survey phase (Objective 1) and controlled evaluation experiment (Objective 3) employ validated instruments, including the NASA-TLX and TAM 2, to generate statistically measurable and replicable results.
2. **Interpretivism:** In contrast, the **qualitative components** of the study, including interviews and focus groups, allow for the exploration of participants' subjective experiences and contextual factors influencing AI tool adoption. Through these means the study adopts an interpretivist view to explore the subjective experiences, workflow nuances, and "Tool–User" contradictions faced by Zambian medical practitioners. This acknowledges that the introduction of AI into a healthcare setting is a social intervention as much as a technical one.

By triangulating these data sources, the research ensures that the findings are not only statistically significant but also valid within the specific constraints of the Zambian public health sector.

3.2 Theoretical Framework: Cultural-Historical Activity Theory (CHAT)

Cultural-Historical Activity Theory provides a lens through which the integration of artificial intelligence (AI) in radiological workflows can be examined in context, not just as a technical advancement, but as a major change within a social, cultural and institutional environment. Originally from the work of Vygotsky [107] and later expanded by Engeström [23], CHAT views human activity as a socially situated, tool-mediated process influenced by cultural norms, historical developments, and institutional structures.

In the context of AI in radiology in Zambia, CHAT helps unpack how radiologists, health systems, AI tools, and broader societal and policy influences interact to shape medical image interpretation.

3.2.1 Core Components of the CHAT Model in this Study

Table 3.1 CHAT elements and their application in the context of radiology in Zambia

CHAT Element	**Application in Radiology-AI Context in Zambia**
Activity	Clinical work associated with medical image interpretation.
Object	Accurate and efficient interpretation of medical images to improve patient outcomes.
Subject	Radiologists, radiographers, and clinicians using or affected by AI tools.
Community	Hospital staff, patients, AI developers, policymakers, and external funders.
Division of Labour	Redistribution of tasks between human actors (radiologists) and non-human agents (AI systems).
Tools/Artifacts	AI diagnostic systems, PACS, and imaging devices.
Rules	Ethical guidelines, medical protocols, AI regulations, and institutional policies.

3.2.2 CHAT as a Lens for Understanding Integration Dynamics

By viewing AI implementation through CHAT, the research recognizes that technological integration is not linear or solely technical. It reveals contradictions and tensions, such as:

- **Tool–User Contradictions**: Radiologists may mistrust AI outputs or lack the training to interpret them effectively.
- **Rules–Object Contradictions**: Regulatory frameworks may lag behind the technology, creating uncertainty in how AI tools can be ethically and legally used.

- **Division of Labor–Community Contradictions**: AI may disrupt existing professional hierarchies or create resistance among radiologists who perceive automation as a threat to their expertise.

These contradictions are not inherently negative; rather, they are opportunities for systemic development and learning. For instance, identifying a contradiction between AI tool capabilities and radiologist expectations may lead to better training programs or interface redesigns that support collaboration instead of replacement.

3.2.3 Why CHAT is Relevant for the Zambian Context

In Zambia, resource limitations, professional shortages, and emerging digital health initiatives present a complex setting in which AI is not just a tool but a potential agent of systemic change. CHAT encourages the researcher to consider:

- How historical and institutional factors (e.g., public health infrastructure, medical education systems) influence AI adoption.
- How cultural attitudes toward automation, authority, and technology shape clinician-AI interaction.
- How social learning processes can support sustainable integration of AI tools in a way that empowers rather than replaces local practitioners.

3.3 Research Design

This research adopted a **complex multi-stage mixed-methods design** structured around the **CRISP-DM (Cross-Industry Standard Process for Data Mining)** lifecycle. This standard industry framework provides a cyclical approach to data mining projects, ensuring that the development of the AI models is firmly rooted in business (clinical) understanding and undergoes rigorous evaluation.

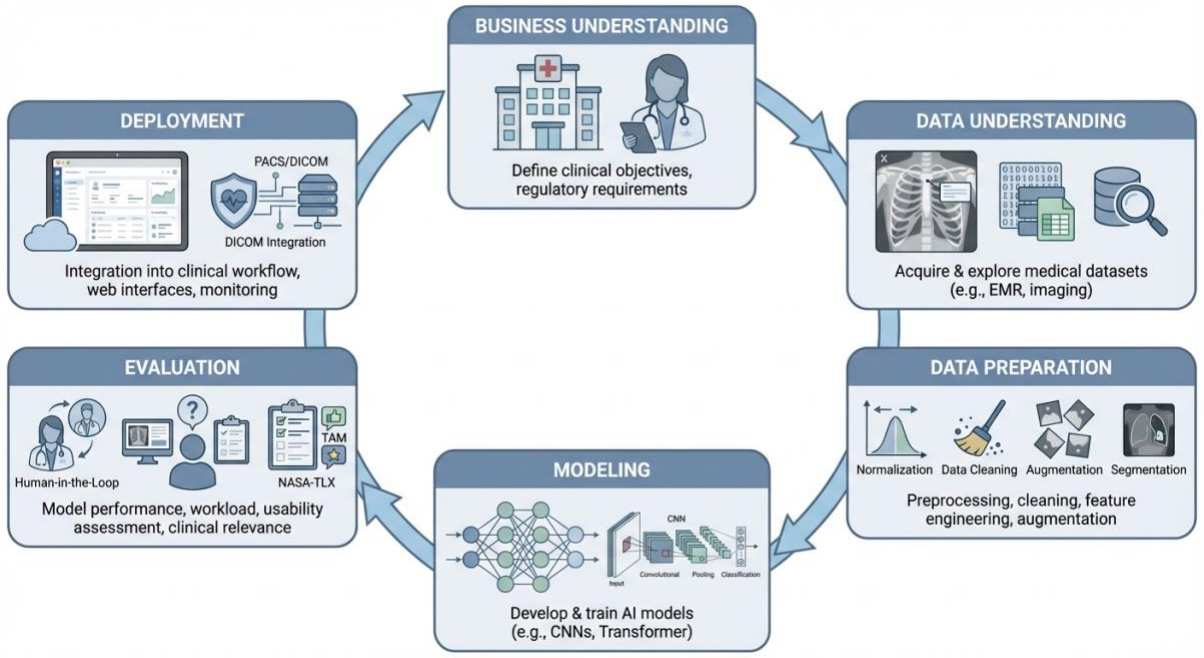


Figure 3.1. CRISP-DM lifecycle as applied to the clinical AI research design, illustrating the six iterative phases from business understanding to deployment. Source: generated with Nano Banana Pro

3.3.1 The CRISP-DM Framework in this Study

The research process was organized into six distinct phases corresponding to the CRISP-DM model, ensuring a logical flow from problem identification to solution evaluation:

1. **Business Understanding (Phase I):** Corresponds to **Objective 1**. This phase involved defining the clinical problem, understanding the Zambian radiological context, and identifying the specific constraints (workload, infrastructure) through surveys and interviews.
2. **Data Understanding (Phase II):** Involved the acquisition and exploration of medical imaging datasets (e.g., RSNA Pneumonia Detection, Paul Mooney CXR) to assess their suitability for the local context.
3. **Data Preparation (Phase III):** Focused on the preprocessing strategies required to make the data suitable for modelling, including resizing, normalization, and augmentation strategies to handle data scarcity.

4. **Modeling (Phase IV):** Corresponds to **Objective 2**. This phase entailed the selection and training of specific Deep Learning architectures (CNN and Object Detection models) to address the identified pathologies.
5. **Evaluation (Phase VI):** Corresponds to **Objective 3**. This phase moved beyond statistical validation to clinical validation, employing controlled experiments (using NASA-TLX and TAM 2) to assess the tool's impact on human subjects.
6. **Deployment (Phase V):** Involved the creation of user interfaces and integration strategies (web-based and DICOM-Reader-integrated) to demonstrate how the models could be operationalized in a real-world setting.

3.3.2 Specific Research Strategies per Objective

The study employed different strategies for each specific research objective:

- **For Objective 1 (Understanding Challenges):** A **descriptive and exploratory strategy** was used. A convergent parallel mixed-methods design was applied, utilizing a cross-sectional survey to quantify workload and demographics, alongside semi-structured interviews to gather qualitative insights into workflow bottlenecks. The quantitative survey data and qualitative interview data were collected concurrently, analyzed independently, and integrated during interpretation to provide a comprehensive understanding of the phenomenon.
- **For Objective 2 (AI Solution Development):** An **experimental and engineering strategy** was adopted. This involved the iterative design and training of machine learning models and the development of a software prototype. The design followed a user-centered approach, incorporating feedback loops from expert practitioners (radiologists) to refine the interface.
- **For Objective 3 (Evaluation):** A **quasi-experimental, within-subjects strategy** was employed. Participants acted as their own controls, performing interpretation tasks under two conditions (with and without AI assistance). This design minimized the impact of individual variability in clinical skill, allowing for a direct comparison of perceived workload and usability between the two workflows.

3.3.3 Time Horizon

The research was primarily **cross-sectional** in nature. The survey (Objective 1) provided a snapshot of the current state of radiological practice in Zambia at a specific point in time. Similarly, the experimental evaluation (Objective 3) measured the immediate impact of the AI tool during a single session. However, the development phase (Objective 2) contained longitudinal elements, as it followed an iterative timeline of system refinement based on sequential focus group feedback.

3.4 Phase I: Business Understanding (Contextual Inquiry)

In alignment with the "Business Understanding" phase of the CRISP-DM model, the initial stage of this research focused on empirically defining the operational realities, constraints, and requirements of the Zambian radiological landscape. This phase was critical to ensuring that the subsequent AI solution (the "modeling" phase) was not developed in a vacuum but was directly responsive to genuine clinical needs. This was in alignment with solving research question one.

3.4.1 Problem Scoping and Target Pathology

To ensure high clinical impact, the scope of the AI intervention was narrowed to focus on **Pneumonia** as the base pathology. This decision was driven by an analysis of national health data, which highlighted pneumonia as a critical burden on the healthcare system.

According to the *Zambian Ministry of Health Annual Health Statistical Report 2017-2021* [120], pneumonia consistently ranks among the top ten leading causes of both morbidity and mortality in the country. Specifically, in 2021, it was identified as the **second highest cause of death**, contributing to **11.4 percent of total recorded deaths** [120]. The rising incidence of pneumonia, coupled with the acute shortage of radiologists identified in the problem statement, creates a severe bottleneck in the diagnostic workflow. Consequently, a tool capable of assisting in the interpretation of chest radiographs for pneumonia provided the most solid grounds for immediate intervention while being a pathology interpreted by a wide range of experts.

3.4.2 Study Settings and Target Population

The studies targeted the public healthcare ecosystem of Zambia, with the survey targeting practitioners from all 3000+ healthcare facilities in Zambia. These facilities were of different ownership types as well as facility levels. The ownership types included the Government of the Republic Of Zambia (GRZ), Private, Non-Governmental Organization (e.g. Mission Hospital) and Military. Whereas the facility levels included the following: Level One, Two, and Three Hospitals, Hospice, Dental Clinic, Diagnostic Centre, Eye Clinic, Fertility Clinic, First-aid stations, General Clinic, Health Centre, Health Post, Mobile Clinic, Optic Clinic, Physiotherapy, Rehabilitation Centre and Specimen collection Centre.

The interviews targeted radiological experts and general doctors from the UTHs and were conducted digitally on Google Meet[6], with the survey using Google Forms.

3.4.3 Quantitative Data Collection: Clinician Survey

To quantify the prevalence of workflow challenges and assess the baseline reception of AI technologies, a cross-sectional survey was administered.

Target Population

The target population of this study comprised **clinicians involved in medical imaging workflows in Zambia** from examination ordering, imaging, through to reporting. Including Medical doctors, Radiologists, Radiographers, Medical Licentiates and Clinical officers working in health facilities in Zambia. To capture a holistic view of the "Division of Labor" (as per the CHAT framework), the inclusion criteria encompassed:

- **Specialists:** Radiologists and Radiology Residents, who represent the expert tier of interpretation.
- **Generalists:** Medical Officers, Medical Licentiates, and other clinicians who, due to the shortage of specialists, frequently interpret images as part of their daily rounds.
- **Allied Health Professionals:** Radiographers and Medical Imagers who are often the first point of contact with the imaging data.

[6] https://workspace.google.com/products/meet/

Inclusion Criteria

Participants were eligible for inclusion if they:

- were qualified medical professionals practicing in Zambia;
- were involved in requesting, performing, or interpreting diagnostic imaging studies;

Sampling Strategy

Nonprobability convenience and snowball sampling was used to reach the disbursed population of medical professionals across the country. Recruitment was conducted primarily through professional digital communication channels, specifically WhatsApp[7] groups. This allowed for the capture of firsthand experiences from the participants. This approach allowed for the rapid dissemination of the survey instrument to practitioners in both urban centres (e.g., Lusaka, Copperbelt) and peripheral facilities. A total of 100 valid responses were obtained and included in the final quantitative analysis.

Instrument Design

Data was collected using a structured electronic questionnaire (hosted on Google Forms[8]). The instrument was designed to capture data across five key dimensions, see Appendix L for the full instrument:

1. **Demographics:** Professional cadre, years of experience, and facility ownership type (Government/Private/Mission).
2. **Experience Profile:** Frequency of interaction with specific imaging modalities (X-ray, CT, MRI).
3. **Workload Assessment:** Likert-scale items measuring perceived burden, study volume, and time pressure.
4. **Operational Effectiveness:** Self-reported confidence in diagnosis and the impact of missing clinical history.
5. **AI Perception and Awareness:** Baseline awareness of Artificial Intelligence and attitudes toward its potential role in healthcare.

[7] https://www.whatsapp.com
[8] https://workspace.google.com/products/forms

Procedure

Prior to full deployment, the questionnaire underwent a pilot phase with a small cohort of medical professionals to validate the clarity and relevance of the items. Following minor refinements, the survey was distributed online. All responses were anonymized, ensuring that participants could report on sensitive issues without fear of professional repercussions. The survey instrument was active for a period of **eight months, from April 2025 to December 2025**, allowing sufficient time for participation across diverse clinical settings.

3.4.4 Qualitative Data Collection: Semi-Structured Interviews

Complementing the quantitative survey, this study employed a **cross-sectional qualitative approach** using semi-structured interviews to explore the specific "tool–user" contradictions and workflow bottlenecks that could not be adequately captured through quantitative measures alone.

Target Population and Inclusion Criteria

The target population for the qualitative component comprised **radiology residents and medical doctors at UTHs**, Participants were eligible for inclusion if they:

1) Had experience carrying out medical image interpretation
2) were currently practicing in a clinical setting within Zambia; and
3) had direct experience with requesting, performing, or interpreting diagnostic imaging studies

Only participants who provided informed consent and expressed willingness to participate in interviews were considered.

Sampling and Recruitment

A **purposive sampling strategy** was employed to select interview participants. Selection prioritized **professional variation**, ensuring representation from both specialist radiology residents and non-specialist clinicians (to capture pragmatic, resource-constrained workflows). A total of **four participants** were recruited and interviewed for the qualitative component of the study.

Interview Protocol

Data collection was guided by a **semi-structured interview protocol** (see Appendix A), which ensured consistency across interviews while allowing flexibility to probe emergent issues. The protocol was organized around three thematic domains:

1. **Workflow Mapping**
   Participants were asked to describe the end-to-end journey of medical images, from acquisition to diagnosis, including tools used (e.g., lightboxes, personal mobile phones) and transfer mechanisms (e.g., physical CDs, messaging applications).
2. **Challenge Identification**
   Probing questions explored workflow pain points such as diagnostic turnaround times, access to expert second opinions, information gaps, and the cognitive workload associated with complex cases.
3. **Requirements Engineering**
   Participants were invited to conceptualize an ideal AI-assisted system, indicating which tasks they would trust to automation (e.g., triage versus definitive diagnosis) and which system features would be critical for clinical adoption.

Data Collection and Analysis Procedure

Interviews were conducted from **February 2024 to September 2025**, via secure video conferencing platforms Google Meet to accommodate clinicians' schedules. All interviews were audio-recorded with explicit participant consent and subsequently transcribed verbatim. This approach ensured analytical rigor and preserved the precise language and contextual nuances of participants' experiences for subsequent thematic analysis.

3.4.5 Data Analysis Strategy for Phase I

In alignment with the convergent mixed-methods design, the quantitative survey data and qualitative interview transcripts were processed independently before the findings were triangulated. This parallel analysis allowed for a comprehensive understanding of the workflow challenges, where statistical trends could be contextualized by the detailed experiences of the practitioners.

Quantitative Survey Analysis

The raw survey data was exported from Google Forms and initially pre-processed using Google Sheets[9]. This stage involved data cleaning and the conversion of Likert scale[10] responses into numerical values to facilitate statistical processing, where "Strongly Disagree" was coded as 1 and "Strongly Agree" was coded as 5.

Subsequent statistical analysis was conducted using JASP[11]. The analysis procedure included:

1. **Descriptive Statistics:** Frequencies and percentages were calculated to summarize respondent demographics (profession, years of practice, facility type) and general perceptions of workload and efficiency.
2. **Inferential Statistics:** Due to the non-normal distribution of the data, non-parametric tests were employed to determine if significant differences existed between groups. The Mann-Whitney U test was used for comparing two independent groups (e.g., gender), while the Kruskal-Wallis H test was utilized for comparisons across three or more groups (e.g., years of practice, professional cadre).

For the presentation of the data a workload mean was used, which was a combination of the workload question results for simpler analysis.

Qualitative Interview Analysis

Audio recordings from the semi-structured interviews were transcribed verbatim using the automated service Turboscribe[12]. To ensure accuracy and maintain confidentiality, the transcripts were reviewed against the original recordings, and all personally identifying information was redacted using Google Docs[13].

[9] https://workspace.google.com/products/sheets/
[10] https://www.scribbr.com/methodology/likert-scale/
[11] https://jasp-stats.org/
[12] https://turboscribe.ai/
[13] https://workspace.google.com/products/docs/

The redacted transcripts were imported into **Taguette**[14], an open-source qualitative data analysis tool, for coding and organization. The analysis followed Braun and Clarke's six-phase framework for reflexive thematic analysis [11]:

1. **Data Familiarisation:** The researcher read and re-read the transcripts to become immersed in the data.
2. **Initial Coding:** Interesting features of the data were systematically identified across the dataset, and initial labels were generated within Taguette.
3. **Generating Themes:** Codes were collated into potential themes, gathering all data relevant to each potential theme.
4. **Reviewing Themes:** The themes were checked against the coded extracts (Level 1) and the entire dataset (Level 2) to ensure they accurately reflected the participants' perspectives.
5. **Defining and Naming Themes:** Ongoing analysis was conducted to refine the specifics of each theme and the overall story the analysis tells.
6. **Producing the Report:** The final themes were selected to provide rich contextual detail that complemented the quantitative survey findings.

To ensure analytical rigor, themes were cross-checked against the raw data to confirm they were well-supported and representative of multiple participants' perspectives rather than isolated anecdotes.

3.5 Phase II & III: Data Understanding and Preparation Strategy

Following the CRISP-DM lifecycle, the research progressed to "Data Understanding" and 'Data Preparation.' Because of the lack of annotated medical imaging data within the Zambian context, this phase required a rigorous selection of proxy datasets that could ensure model viability while approximating the clinical targets identified during the contextual inquiry.

[14] https://www.taguette.org/

3.5.1 Dataset Selection Criteria

The selection of appropriate training data was governed by a set of clinical and technical criteria established in consultation with an expert radiologist. These criteria prioritized:

- **Relevance:** The data must represent the pathology identified as a high-priority target (Pneumonia) in a modality widely accessible in Zambia (Chest X-Ray).
- **Orientation:** Images must be in standard Anterior-Posterior (AP) or Posterior-Anterior (PA) views, consistent with routine clinical protocols.
- **Volume:** The dataset needed to be sufficiently large to train Deep Learning models.
- **Labels:** Reliability of ground-truth annotations was paramount for supervised learning.
- **Literary reference:** The dataset should preferably be cited by publications.

Based on these criteria, two primary public datasets were selected:

1. **Paul Mooney Chest X-Ray Dataset** [51]**:** Selected for the **classification** task. This dataset is widely cited in literature and provides a substantial volume (5,856 images) of categorized images (Normal vs. Pneumonia), making it suitable for training a binary classifier. It is the most popular CXR dataset on Kaggle[15]
2. **RSNA Pneumonia Detection Challenge Dataset** [85]**:** Selected for the **detection** task. Unlike standard classification datasets, this source includes granular bounding-box annotations identifying the specific regions of lung opacity. This level of detail was essential for developing the "explainable" localization feature requested by clinicians during the contextual inquiry.

3.5.2 Strategy for Data Preparation and Augmentation

Medical imaging data inherently presents challenges regarding class imbalance and variability in image quality. To prepare the raw data for modelling, a systematic preprocessing pipeline was designed.

[15] https://www.kaggle.com/

Preprocessing Strategy

Standardization was applied to ensure consistent input for the neural networks. This involved:

- **Resizing:** Downscaling high-resolution DICOM/JPEG images to standard dimensions compatible with the chosen network architectures (e.g., VGG16 input requirements) to balance computational efficiency with feature preservation. Aspect ratio was preserved.
- **Normalization:** Scaling pixel intensity values (typically 0-255) to a normalized range (0-1) to facilitate faster convergence during the training process.

Data Augmentation Strategy

A critical challenge in applying Deep Learning to medical imaging is the risk of overfitting, particularly given the relatively small size of the datasets compared to natural image datasets. To address this, a **Data Augmentation Strategy** was employed during the training phase.

The strategy focused on generating synthetic variations of the training images to simulate real-world clinical scenarios encountered in Zambian facilities:

- **Geometric Transformations:** Random rotations, shears, and translations were applied to mimic variations in patient positioning and alignment. For the detection task, this strategy included a constraint to ensure that bounding box annotations were transformed synchronously with the image to maintain ground-truth integrity.
- **Photometric Transformations:** Adjustments to brightness and contrast were applied to simulate the diversity of exposure settings found across different X-ray machines, ranging from modern digital detectors to older analogue units.

Data Splitting Protocol

To ensure a rigorous evaluation, holdout split was enforced across all modelling tasks. Datasets were partitioned into three distinct subsets: a **Training Set** for learning parameters, a **Validation Set** for hyperparameter tuning and early stopping, and a **Test Set** held out exclusively for the final performance evaluation. This separation ensured that reported metrics reflected the model's ability to generalize to unseen patient data.

3.6 Phase IV: Modelling Strategy

This phase focused on the selection and development of machine learning architectures capable of addressing the clinical needs identified in Phase I. Two distinct modelling approaches were pursued: a **binary classification** model to answer the 'what' (presence of disease) and an **object detection** model to answer the "where" (localization of disease). This dual strategy was designed to provide a comprehensive diagnostic aid that supports both triage efficiency and diagnostic explainability.

3.6.1 Rationale for Transfer Learning (VGG16)

For the task of binary classification (Normal vs. Pneumonia), a **Transfer Learning** approach was adopted. Training deep CNNs from scratch requires massive datasets and significant computational resources, both of which are constraints in this context. Transfer learning mitigates this by leveraging knowledge from a pre-trained model.

The **VGG16 architecture** [93] was selected as the backbone for the classification model. Pre-trained on the ImageNet [18] dataset, VGG16 has demonstrated robust feature extraction capabilities across a wide range of visual tasks. Its architecture, characterized by a uniform arrangement of convolutional and pooling layers, is well-suited for capturing the hierarchical features present in medical images, from low-level edges (ribs, diaphragm) to high-level patterns (consolidations). By freezing the early convolutional layers and retraining only the final classification layers, the model could effectively learn to distinguish pneumonia features without requiring a dataset the size of ImageNet.

3.6.2 Rationale for Object Detection

To address the qualitative finding that clinicians require 'explainable' AI (i.e., knowing *why* a decision was made), an object detection model was implemented to localize suspected regions of pathology.

The **Mask R-CNN (Region-based Convolutional Neural Network)** [34] architecture was chosen for this task. Mask R-CNN is a state-of-the-art instance segmentation framework that extends Faster R-CNN. It was selected for its ability to simultaneously perform object detection (drawing bounding boxes around lesions) and generate segmentation masks. This capability is

particularly valuable in medical imaging, where pathologies like pneumonia often present as diffuse opacities rather than distinct objects. The use of a ResNet-50 [33] backbone with a Feature Pyramid Network (FPN) ensured the model could detect lesions of varying sizes, addressing the clinical reality that pneumonia can manifest as both small focal patches and large lobar consolidations.

3.6.3 Training Protocols

To maximize model performance and ensure stability during the learning process, specific training protocols were defined:

Phased Unfreezing Strategy

For the VGG16 model, a fine-tuning strategy was employed. Initially, the heavy convolutional base was "frozen," preventing its weights from being updated. Only the newly added custom classifier layers were trained. This allowed the new layers to learn the specific mapping for pneumonia classification without destroying the pre-learned feature extractors. Subsequently, a fine-tuning phase was conducted where the top layers of the VGG16 base were 'unfrozen' and trained with a very low learning rate, allowing the model to adapt its high-level feature representations specifically to chest X-rays.

Optimization and Loss Functions

- The **Adam optimizer** [53] was selected for its adaptive learning rate capabilities, which facilitate efficient convergence.
- **Binary Cross-Entropy Loss** [10] was used for the classification task, appropriate for the binary nature of the diagnosis.
- **Multi-task Loss** (combining classification, bounding box coordinates, and mask loss) was inherent to the Mask R-CNN training, ensuring the model simultaneously learned to identify, locate, and delineate the pathology [34].

Performance Monitoring

Training progress was rigorously monitored using validation sets. Metrics such as validation loss and accuracy were tracked at each epoch. An 'Early Stopping' mechanism was

implemented as a heuristic to halt training when validation performance ceased to improve, thereby preventing the model from memorizing the training data (overfitting) and ensuring better generalization to new patient images.

3.7 Phase V: Deployment Strategy (Iterative Design)

The deployment phase focused on bridging the gap between algorithmic performance and clinical utility. Following the principles of **User-Centered Design**, the development of the interface was not a linear software engineering task but an iterative cycle of prototyping, evaluation, and refinement. This strategy was essential to address the 'Tool-User' contradictions identified in the CHAT framework, ensuring that the final artifact aligned with the existing workflows and mental models of Zambian practitioners.

3.7.1 The User-Centered Design Process

The design process was structured to facilitate continuous interaction between the developer and the target users. Instead of delivering a final product at the end of the research, a functional prototype was introduced early to elicit concrete feedback. This process involved three key stages:

1. **Initial Prototyping:** Development of a 'minimum viable product' capable of the core function: accepting a DICOM image and displaying the AI model's output. This version prioritized functional transparency over feature density.
2. **Iterative Refinement:** Modification of the interface based on direct expert feedback. This involved changing not just visual elements (e.g., colour of bounding boxes) but also workflow logic (e.g., automating the inference trigger upon upload).
3. **Contextual Adaptation:** Expansion of the tool's capabilities to reflect the complexity of real-world diagnosis, moving from single-pathology detection to a multi-pathology display framework in response to clinician requests.

3.7.2 Focus Group Methodology

To guide these iterations, qualitative data was collected through sequential focus group discussions. These sessions served as "diagnostic checkpoints" for the technology.

Participants

The focus groups comprised a targeted mix of domain experts, including a qualified Radiologist and a Radiology Resident. This composition ensured that feedback captured both high-level diagnostic requirements and the practical, day-to-day workflow realities faced by trainees. Including an expert in HCI in the first session provided an additional layer of scrutiny regarding usability standards.

Procedure

The focus group sessions were structured as interactive walkthroughs.

1. **First Session (Post-Prototype):** Participants were tasked with using the initial interface to interpret a set of test images. The discussion focused on "friction points" aspects of the tool that slowed down interpretation or caused confusion. Key topics included the clarity of the AI's "opinion" (e.g., probability scores vs. binary labels) and the visualization of the bounding boxes.
2. **Second Session (Post-Refinement):** Following the implementation of changes, a second session was held to validate the improvements. This session shifted the focus from basic usability to clinical relevance. The discussion probed whether the tool's output, now refined, actually aided diagnostic reasoning. It was during this phase that the requirement for "multi-pathology" support emerged, highlighting the practitioners' need for a holistic diagnostic aid rather than a single-disease detector.

The insights gathered from these sessions were systematically coded and translated into technical requirements (detailed in Chapter 4), ensuring that the final interface used in the evaluation phase was a co-created artifact reflecting the specific needs of the local user base.

Table 3.2 Development Order of the AI Interpretation Interface

Stage	Description	Key Outcomes
1. Initial Development	Built the interface in Next.js 13.4 with three modes: Classification Tool, Detection Tool, and Combined Tool. Added drag-and-drop DICOM upload and single-click inference.	Functional prototype enabling AI-assisted pneumonia interpretation.
2. First Focus Group	Discussion with 1 qualified radiologist and 1 radiology resident on usability and clinical relevance.	Suggestions implemented to improve workflow integration and visual clarity.
3. Evaluation Phase	Conducted usefulness, usability, and workload impact study. Used counterbalanced task order (AI-assisted first vs. standard tool first) with NASA-TLX and TAM 2 questionnaires.	Quantitative and qualitative data collected on tool performance and perceived value.
4. Second Focus Group	Discussion with the same experts after partial evaluation results. Focus shifted towards multi-pathology interpretation.	Interface updated to display multiple suspected pathologies with enhanced results layout and usability improvements.
5. Full Experiment in Interface	A version of the interface was adjusted to allow the carrying out of the usefulness experiment digitally	The interface was updated to allow for the full experiment data collection.

3.8 Phase VI: Evaluation Methodology

The final phase of the research aimed to empirically validate the impact of the developed AI solution on the clinical workflow. Moving beyond algorithmic performance metrics (such as accuracy or F1-score), this evaluation focused on the human-computer interaction aspect, specifically assessing perceived workload and user acceptance. A rigorously controlled

experimental design was employed to isolate the effect of the AI intervention from individual practitioner variability.

3.8.1 Experimental Design

A **quasi-experimental, within-subjects (repeated measures) design** was adopted. This design was chosen because of the scarcity of expertise on the subject and for its statistical efficiency; by having each participant act as their own control, the study could effectively control for the wide variance in individual clinical experience and diagnostic skill levels found in the target population. This experiment was cross-sectional.

Target Population and Inclusion Criteria

The target population of this study was medical doctors practicing in Zambia, with a bias towards radiological experts like radiological residents.

Participants were eligible for inclusion if they:

- were qualified medical doctors practicing in Zambia;
- were involved in requesting or interpreting diagnostic imaging studies;

Sampling Strategy and recruitment

Non probability convenience [99] and snowball sampling [99] was used to reach the medical doctors involved in the study. Recruitment of participants was conducted manually by physically visiting the UTHs Adult Hospital Radiology department reporting room and scheduling time with the radiology residents. Once the residents were exhausted they were asked to recommend colleagues they felt would best participate in the study. The study was then conducted with or shared with those medical doctors. A total of 47 participants were interacted with for this study. **The interactions ran from February 2024 to November 2025**

Counterbalancing

To mitigate potential order effects such as practice effects (improving due to repetition) or fatigue effects (performance dropping over time), the task order was **counterbalanced**. Participants were randomly assigned to one of two sequences:

- **Group A (Those whose participant numbers were even):** Performed the interpretation task *with* AI assistance first, followed by the task *without* assistance.
- **Group B (Those whose participant numbers were odd):** Performed the interpretation task *without* AI assistance first, followed by the task *with* assistance.

The interpretation task itself remained consistent: reviewing a chest radiograph suspected of pathology and making a determination regarding the presence and localization of disease. This rigorous structure allowed for a direct, unbiased comparison of the two workflows.

3.8.2 Experimental Setting and Procedure

The experimental evaluation was conducted using a controlled clinical simulation designed to replicate routine chest radiograph interpretation workflows while minimizing variability arising from system performance and network conditions. The study evaluated two interpretation workflows: a standard non–AI-assisted workflow and an AI-assisted workflow.

To ensure consistency in system responsiveness and to isolate the cognitive impact of AI assistance from technical variability, the interaction between the frontend interface and the AI inference process was deliberately controlled. Specifically, AI outputs were pre-generated and stored within the interface, and a fixed **one-second artificial delay** was introduced before displaying AI results to participants. This simulated a realistic inference latency while ensuring that all participants experienced identical response times. This approach eliminated fluctuations due to network instability, server load, or device-specific connectivity differences, thereby standardizing the perceived inference time across all experimental sessions.

For the **non–AI-assisted workflow**, participants interpreted chest radiographs using the **Orthanc DICOM Viewer**, reflecting conventional clinical image review practices without algorithmic support.

The experiment was conducted in two settings:

- **On-site sessions**, involving radiologists, radiology residents and medical doctors from the UTHs in Lusaka, Zambia, were carried out on researcher-controlled computers configured identically in terms of software, display settings, and input devices.

- **Remote sessions**, involving participants from other institutions, were conducted over the internet using participants' own devices. These participants accessed the same experimental interface and datasets, with identical system controls applied, including the fixed AI response delay and standardized image presentation, to maintain procedural consistency.

The experimental procedure followed a standardized sequence for all participants. First, participants completed an informed consent form and a biodata questionnaire capturing demographic and professional information. Following this, participants proceeded to the interpretation tasks according to the counterbalanced workflow order assigned to them (AI-assisted first or non–AI-assisted first).

In each workflow condition, participants reviewed the same chest radiograph(s) suspected of pathology and were instructed to determine the presence and, where applicable, the anatomical localization of disease. Participants were encouraged to perform the task as they would under normal clinical circumstances. No diagnostic feedback was provided during or between conditions to prevent learning effects.

Immediately after completing each workflow condition, participants completed the NASA Task Load Index (NASA-TLX) questionnaire to assess perceived workload associated with that specific workflow. Upon completion of the AI-assisted condition, participants were administered the Technology Acceptance Model 2 (TAM 2) questionnaire to evaluate overall perceptions of usefulness, ease of use, and adoption intent regarding the AI-assisted system.

All experimental sessions adhered strictly to this protocol, ensuring consistency across participants, settings, and workflow conditions.

3.8.3 Instruments

Two standardized, validated psychometric instruments were employed to capture quantitative data regarding the user experience. Both tools were implemented in Google Forms for simpler data processing, and analysis.

NASA Task Load Index (NASA-TLX)

To measure the perceived workload, the **NASA-TLX** [31] was utilized. This multi-dimensional assessment tool rates workload across six distinct subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. By collecting NASA-TLX scores immediately after each condition (Standard vs. AI-Assisted), the study could quantify not just the *amount* of work, but the *nature* of the cognitive burden imposed or alleviated by the AI tool.

The use of the NASA-TLX was slightly modified to not use the weighting (Raw). Raw NASA-TLX results have been shown to have high correlation with weighted results [29, 87]. This was done to shorten the length of the experiments due to the high level of unavailability of experts in radiological image interpretation.

Technology Acceptance Model 2 (TAM 2)

To predict the likelihood of long-term adoption, a modified version of the **TAM 2 questionnaire** [105] was administered. The TAM 2 framework posits that *Perceived Usefulness (PU)* and *Perceived Ease of Use (PEOU)* are the primary determinants of a user's intention to use a new system. Additional constructs relevant to the professional setting, such as "Job Relevance" and "Output Quality," were included to provide a granular understanding of how the tool fits into the professional identity and requirements of the medical staff.

3.8.4 Data Analysis Procedure

The results from both instruments were copied from Google Forms to Google Sheets where preprocessing took place. All quantitative data from the experimental evaluation were then processed and analyzed using **JASP**. The analysis was divided into two distinct components: determining the overall impact of the tool on workload and usability, and investigating how these effects varied across different demographic groups.

Data Preprocessing and Variable Operationalization

Prior to analysis, the raw data collected from the NASA-TLX and TAM 2 instruments were pre-processed:

- **Workload Index Construction:** To facilitate analysis, an attempt was made to combine the workload-related items (‘volume affects quality’, ‘too many studies per day’ and 'experience time pressure') into an Aggregate Workload Index.
- **Dependent Variable Calculation (NASA_DIFF):** To measure the specific impact of the AI intervention, a difference score (NASA_DIFF) was computed for each participant. This was defined as the **Standard Tool Workload score minus the AI-Assisted Tool Workload score**. A positive NASA_DIFF indicated a reduction in perceived workload when using the AI tool.
- **TAM 2 Scoring:** The numerically coded responses for the Technology Acceptance Model (TAM 2) constructs (Perceived Usefulness, Ease of Use, Intention to Use) were aggregated from the 7-point Likert scale items to derive mean composite scores for each construct.

Statistical Tests and Justification

Descriptive statistics (Mean, Standard Deviation) were generated to provide a baseline understanding of the perceived workload and user acceptance. For inferential analysis, tests for normality (Shapiro-Wilk [90]) indicated that the data were not normally distributed. Consequently, non-parametric statistical tests were employed to ensure robustness:

1. **Mann-Whitney U Test:** This test was utilized to compare differences between two independent groups. Specifically, it was used to determine if there were statistically significant differences in NASA_DIFF (workload reduction) and TAM 2 construct scores between **genders** (Male vs. Female).
2. **Kruskal-Wallis H Test:** This test was applied to compare differences across three or more independent groups. It was used to analyze the influence of **Professional Cadre** (e.g., Medical Doctor vs. Radiologist vs. Medical Imager) and **Years of Practice** on the perceived workload reduction and tool usefulness.
3. **Post-Hoc Analysis:** In instances where the Kruskal-Wallis test yielded a statistically significant result, **Dunn’s post-hoc tests** [20] with **Holm correction** [37] were conducted to identify exactly which pairs of groups differed significantly while controlling for the family-wise error rate.

4. Preliminary analysis of the participant distribution revealed low counts for specific cadres (e.g., Radiologists, n=1). To ensure sufficient statistical power and avoid invalid comparisons, professional cadres were collapsed into two functional groups for analysis: **"Specialists"** (comprising Radiologists and Radiology Residents) and **"Generalists"** (comprising Medical Doctors and other clinical officers).

All statistical tests were two-tailed, and a significance level (alpha) of **$p < .05$** was established as the threshold for statistical significance.

3.9. Ethical Considerations

These studies involve contact with individuals' medical records, data that is considered extremely sensitive. This meant a need for rigorous ethical standards, including anonymity as it relates to surveys and interviews that involve humans. Ethical clearance was obtained from the following authorities:

University of Zambia Biomedical Research Ethics Committee Approval Number : REF.2731-2022 see Appendix F renewed in 2023 (see Appendix N) and again in 2025 (see Appendix O). National Health Research Authority Ethical Clearance Approval Reference Number: NRHA00024/10/05/2022 (see Appendix G). Republic of Zambia Ministry of Health Ethical Clearance Approval obtained on 5th September 2022 (see Appendix H). University Teaching Hospitals Ethical Approval (see Appendix I & J). Levy Mwanawasa University Teaching Hospital Approval obtained on 14th June 2022 (see Appendix K). An application was also made with the Natural and Applied Sciences Research Ethics Committee (NASREC) and is pending.

Chapter 4: System Design and Implementation

This chapter presents the technical realization of the proposed AI-assisted radiological interpretation system, translating the research objectives defined in Chapter One into a functional software artefact. While the preceding chapter detailed the methodological framework and evaluation strategies, this chapter focuses on the design, implementation, and deployment of the engineered solution itself.

The system was developed to address the practical constraints identified in the Zambian healthcare context, specifically high radiological workloads, limited access to specialist expertise, and constrained computational infrastructure. Consequently, the design emphasizes modularity, computational efficiency, and seamless integration into existing clinical workflows.

To provide a coherent technical narrative, this chapter is structured to follow the flow of data through the system:

- **User-Driven Design Requirements (Section 4.1):** Requirements for system design from users.
- **System Architecture (Section 4.2):** High-level design of the distributed client–server infrastructure.
- **Data Pipeline (Section 4.3):** Ingestion, normalization, and preprocessing of medical images.
- **AI Model Implementation (Section 4.4):** Architectural configuration and fine-tuning of the deep learning models.
- **Application Interface (Section 4.5):** Development of the clinician-facing web interface.
- **Server-Side Logic (Section 4.6):** Application programming interface (API) design and inference orchestration.
- **Clinical Integrations (Section 4.7):** Extension into DICOM viewers and browser-based workflows.

4.1 User-Driven Design Requirements

To ensure the developed system addressed the actual operational constraints of the Zambian context, the technical specifications were directly derived from the 'Specific Requests for AI Automation' elicited during the clinician interviews (Phase I). As identified in the contextual inquiry, practitioners rejected 'black box' automation in favour of a supportive, explainable tool. The participant profiles are described in Table 5.4 These qualitative insights were translated into the following core technical requirements:

1. Requirement for Localization and Quantification

Participants explicitly requested features that go beyond simple binary diagnosis. A [RaRe] participant emphasized the need for tools that manage 'trivial' yet time-consuming tasks and provide quality assurance, stating: *'*Give me the dimension of a particular pathology... tell me the quality of the image... and highlight the areas that I'm supposed to pay particular attention to.'

- **Technical Translation:** This necessitated the implementation of an **Object Detection and Segmentation model (Mask R-CNN)** rather than a simple classifier. The system was engineered to generate bounding boxes and segmentation masks to visually 'highlight' specific regions of interest (ROI), satisfying the need for spatial explainability.

2. Requirement for 'Spark Plug' Diagnostic Suggestions

Generalist practitioners, such as a [SRMO], conceptualized the AI not as a replacement but as a cognitive primer. The SRMO requested a tool that offers 'suggestions of what might be present,' acting as a 'spark plug' or 'just a hint or a hunch to get you started.'

- **Technical Translation:** This drove the design of the **Multi-Pathology Dashboard** (Interface v3.0). Instead of a definitive single output, the interface was designed to present a ranked probability list of suspected conditions (e.g., Pneumonia vs. Tuberculosis), functioning as a differential diagnosis prompt to trigger the clinician's own reasoning process.

3. Requirement for Triage Prioritization

The interviews pinpointed crucial targets for an interpretation tool: emergencies (like ‘air under the diaphragm,’ ‘pneumothorax’), trauma (‘fractures’), and common high-burden pathologies (*‘pneumonias’*).

- **Technical Translation:** This informed the **Model Selection Strategy**. Priority was given to training models on pathologies that require immediate intervention (Pneumonia) to serve as an effective triage mechanism in high-volume, acute care settings.

4.2 System Architecture

The system adopts a **modular, distributed client–server architecture**, deliberately separating user interaction from computationally intensive machine learning inference tasks. This architectural choice ensures that clinician-facing applications remain lightweight and responsive, even when deployed on lower-specification hardware typical of resource-constrained clinical environments, while centralized server infrastructure handles the heavy computational workload.

By decoupling the frontend and backend components, the architecture enables independent scaling, updating, and maintenance of system components. This design is particularly well-suited to healthcare settings in the Global South, where minimizing on-site maintenance requirements and reducing software update complexity are critical factors influencing technology adoption.

The architecture is organized into three logical layers:

1. **Presentation Layer (Client)**
2. **Service Orchestration Layer (API)**
3. **Inference Layer (AI Models)**

This layered design also allows for the extension of the tool’s usefulness by integration of external services with the API layer directly.

4.2.1 Architectural Components and Interaction Flow

Communication between system components is implemented using RESTful APIs[16], enabling interoperability across diverse client interfaces.

Presentation Layer

The client-side ecosystem supports multiple entry points into the radiological workflow. The primary interface is a web-based application developed using the Next.js framework (version 13.4)[17], selected for its server-side rendering capabilities and efficient routing under low-bandwidth conditions. The client enables DICOM file ingestion, metadata extraction, and visualization of AI-generated diagnostic outputs.

To support real-world clinical adoption, integration modules were also developed for the Orthanc DICOM server[18] and standard web browsers, allowing the system to embed directly into existing PACS workflows.

The client layer supports multiple entry points into the radiological workflow:

- A standalone web application accessible via standard browsers.
- Integration modules for the Orthanc DICOM server, enabling embedding within existing PACS environments.

Client-side responsibilities include:

- DICOM file ingestion and upload
- Extraction and transmission of relevant metadata
- Rendering of medical images
- Visualization of diagnostic outputs, such as classification results and bounding box overlays

[16] https://www.ibm.com/think/topics/rest-apis
[17] https://nextjs.org/
[18] https://www.orthanc-server.com/

This design allows clinicians to interact with the system using familiar tools, minimizing workflow disruption.

Service Orchestration Layer

The backend is implemented using Python[19] and the Flask micro-framework[20], functioning as the system's orchestration layer. It exposes task-specific endpoints (e.g., /predict for classification and /get_boxes for detection), handling request validation, data serialization, and routing to the appropriate inference model. It also has some endpoints to handle direct interactions with the image like converting from DICOM to PNG or to draw the results directly to the image.

Upon receiving a request, the API server performs:

- Input validation and serialization
- Temporary storage and memory management
- Routing of data to the appropriate preprocessing and inference pipelines

This layer supports concurrent requests, enabling multiple clinicians to interact with the system simultaneously without blocking inference execution.

Inference Layer

The Inference Layer constitutes the computational core of the system. It hosts two deep learning models: a **VGG16-based convolutional neural network** for binary pneumonia classification and a **Mask R-CNN model** for lesion localization and visualization.

Both models are pre-loaded into memory to minimize inference latency and are executed within a GPU-accelerated environment. The separation of inference logic from the API layer ensures that updates to model weights, thresholds, or architectures can be deployed centrally without requiring changes to client applications.

[19] https://www.python.org/
[20] https://flask.palletsprojects.com/en/stable/

The interaction between the Presentation Layer, Service Orchestration Layer, and Inference Layer is illustrated in **Figure 4.1**.

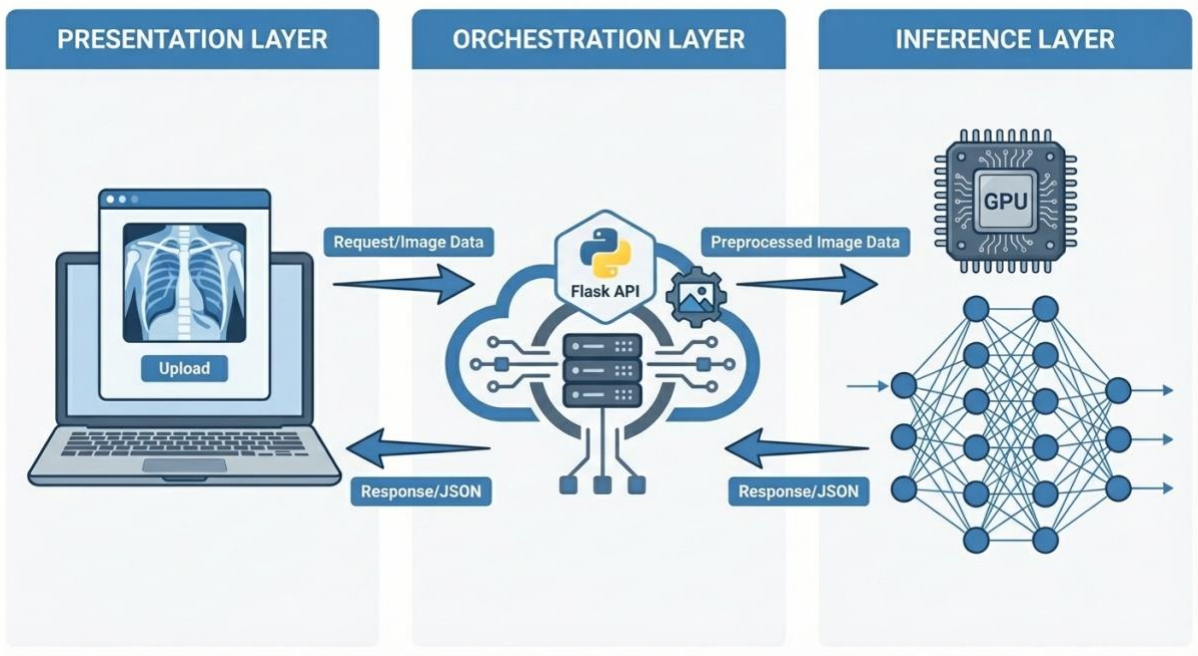


Figure 4.1 A high-level system diagram showing the layer interactions. Source: Generated with Nano Banana Pro

4.2.2 Deployment Environment

To emulate a realistic deployment scenario, the backend infrastructure was hosted on a centralized Microsoft Azure[21] cloud instance running Windows 11 Professional[22] . GPU acceleration was enabled to optimize inference performance. This centralized 'hub-and-spoke' deployment model allows a single high-performance server to support multiple peripheral clinics, mitigating local hardware limitations common in rural and under-resourced facilities.

Importantly, this architecture enables centralized maintenance: updates to AI models or inference logic can be deployed server-side without requiring manual updates on clinician workstations, addressing a key barrier to technology adoption in Global South healthcare systems.

[21] https://azure.microsoft.com/en-us
[22] https://www.microsoft.com/en-us/software-download/windows11

4.3 Data Pipeline Implementation

The efficacy of Deep Learning models in medical imaging is predicated on the quality and consistency of the input data. To ensure the models received standardized tensors capable of facilitating convergence, a robust data pipeline was implemented. This pipeline is divided into two distinct processing stages: **Preprocessing**, which standardizes the raw DICOM input, and **Augmentation**, which applies a stochastic heuristic to enhance model robustness against **clinical** variations.

4.3.1 Image Preprocessing and Normalization Logic

The system utilizes the Pydicom[23] library to parse incoming .dcm files. Pixel data is extracted as a NumPy array. Since chest radiographs typically possess a high dynamic range (12-bit or 16-bit depth), the pixel intensities are normalized using Min-Max scaling to a floating-point range of [0, 1]. This normalization is critical for ensuring that the model's gradients remain stable during backpropagation.

Dimensionality and Channel Adaptation

Both VGG16 and ResNet-50 architectures were originally trained on ImageNet, which consists of 3-channel RGB images. To maintain compatibility with these pre-trained weights:

1. **Channel Replication:** The single-channel grayscale radiograph is stacked along the depth axis to create a pseudo-RGB image (H, W, 3).
2. **Resizing:** The spatial dimensions are adjusted based on the specific model requirements:
    - **Classification Pipeline:** Images are resized to 150 * 150 pixels to reduce the total parameter count in the fully connected layers, optimizing for computational speed.
    - **Detection Pipeline:** Images are resized to 256 * 256 pixels (during inference) or 1024 * 1024 (during training) to preserve fine-grained spatial features necessary for accurate bounding box regression.

[23] https://pydicom.github.io/

4.3.2 The Data Augmentation Heuristic

To mitigate the risk of overfitting, a prevalent challenge when training on finite medical datasets, and to ensure the model generalizes well to the varying acquisition conditions found in Zambian facilities, a **Controlled Robustness Heuristic** was developed.

This heuristic is not merely a collection of random transformations but a structured procedure designed to simulate specific physical and optical imperfections common in low-resource radiological settings (e.g., patient misalignment, older X-ray tubes, or poor focus). The process is divided into three logical phases, detailed below, with the formal algorithmic implementation provided in **Algorithm 4.1**.

Phase 1: Geometric Affine Transformations

The first phase applies global geometric distortions to simulate slight camera misalignments and variations in patient positioning. Parameters are sampled uniformly from conservative bounds to ensure anatomical features remain centred and recognizable:

- **Scaling:** To account for variations in the patient-to-detector distance, the image is randomly rescaled by a factor $s \sim U(1-\delta_{scale}, 1+\delta_{scale})$, resulting in $X_s = S_s(X)$.
- **Translation:** To mimic framing errors, the image is translated by offsets $t_v \sim U(-\delta_v, +\delta_v)$, $t_h \sim U(-\delta_h, +\delta_h)$, where $\delta_v = 0.04H$ and $\delta_h = 0.02W$. The transformed image is given by $X_t = T_{(th,tv)}(X_s)$.
- **Rotation and Shear:** Rotational noise is applied with $\theta \sim U(-\delta_{rot}, +\delta_{rot})$, accompanied by a minor shear transformation $\phi \sim U(-\delta_{shear}, +\delta_{shear})$, yielding the affine-transformed image $X_g = A_{(\theta,\phi)}(X_t)$.

Phase 2: Photometric Jitter

Radiographs captured on analog film or older computed radiography (CR) systems often exhibit inconsistent exposure. This phase introduces controlled non-linear intensity variations.

- **Contrast and Brightness:** The pixel intensities are perturbed by applying random brightness (β) and contrast (γ) factors, effectively simulating over-exposed or under-exposed acquisitions.

Phase 3: High-Frequency Injection

The final phase addresses image quality degradation caused by focus issues or patient movement. A probabilistic branch is executed where the image is either blurred (using a Gaussian kernel with σ) to simulate motion blur, or sharpened (using an unsharp masking factor α) to improve edge definition in lower-quality scans.

Algorithm 4.1. Controlled Robustness Heuristic

Input: Source Image I, Dimensions H (Height in pixels), W (Width in pixels), C (Color Channels)
Output: Augmented Image I'
Hyperparameters:

Δ_scale = 0.02 // Max scale distortion (±2% of original image dimensions)
Δ_rot = 2.0 // Max rotation angle (±2 degrees)
Δ_shear = 1.0 // Max shear angle (±1 degree)
Δ_v = 0.04 // Max vertical shift (±4% of Height H, calculated in pixels)
δ_h = 0.02 // Max horizontal shift (±2% of Width W, calculated in pixels)
δ_photo = 0.1 // Max photometric shift (±10% adjustment on a 0.0 to 1.0 pixel intensity scale)
σ_max = 0.1 // Max Gaussian blur standard deviation (kernel size factor)
α_max = 0.1 // Max sharpening intensity factor

procedure AUGMENT(I, H, W)
// Phase 1: Geometric Transforms (Modifies spatial pixel coordinates)
s ← Uniform(1 - δ_scale, 1 + δ_scale) // Sample scale multiplier (0.98x to 1.02x)
t_v ← Uniform(-δ_v * H, +δ_v * H) // Sample vertical translation in pixels
t_h ← Uniform(-δ_h * W, +δ_h * W) // Sample horizontal translation in pixels
θ ← Uniform(-δ_rot, +δ_rot) // Sample rotation angle in degrees
φ ← Uniform(-δ_shear, +δ_shear) // Sample shear angle in degrees
I_geo ← AffineTransform(I, {s, t_v, t_h, θ, φ})// Apply sampled geometric shifts simultaneously

9: // Phase 2: Photometric Transforms (Modifies pixel intensity values)
10: β ← Uniform(-δ_photo, +δ_photo) // Brightness
11: γ ← Uniform(1 - δ_photo, 1 + δ_photo) // Sample contrast multiplier (0.9x to 1.1x)
12: I_photo ← Clip(I_geo * γ + β, 0, 1) // Apply intensity shifts & clip to valid [0, 1] range

13: // Phase 3: Image Frequency Transformations (Simulates equipment focus variations)
14: p ← Uniform(0, 1) // Generate probability p to pick blur vs. sharpen
15: **if** p < 0.5 **then**
16: σ ← Uniform(0, σ_max) // Sample blur severity
17: I' ← GaussianBlur(I_photo, σ) // Apply Gaussian blur (simulates poor focus/motion)
18: **else**
19: α ← Uniform(0, α_max) // Sample sharpening severity
20: I' ← Sharpen(I_photo, α) // Apply sharpening filter (simulates edge-enhancement)
21: **end if**

22: **return** I'
23: **end procedure**

4.4 AI Model Implementation

This section details the architectural specifications and training methodologies for the two Deep Learning models developed in this study. The implementation strategy prioritized the use of established, pre-trained architectures, specifically VGG16 for classification and Mask R-CNN for object detection, optimized via transfer learning to perform effectively on the specialized task of pneumonia identification.

4.4.1 Pneumonia Classification: VGG16 Architecture

For the binary classification task (determining the presence or absence of pneumonia), a CNN based on the **VGG16** architecture [93] was implemented.

Architectural Modifications

The standard VGG16 model, originally designed for 1,000-class classification on ImageNet, was adapted for binary medical diagnostics. The implementation involved loading the convolutional base (include_top=False) to leverage its robust feature extraction capabilities while replacing the original dense layers with a custom, lightweight classifier head. This modification reduces the parameter count and mitigates overfitting on the smaller medical dataset.

The specific layer configuration is detailed in **Table 4.1**. The convolutional base serves as a fixed feature extractor during the initial training phase, while the custom head, consisting of a dense layer with ReLU activation and a high dropout rate, learns the non-linear mappings specific to pneumonia pathology.

Table 4.1 Detailed layer specification of the VGG16-based architecture

Layer (Type)	**Output Shape**	**Params #**	**Configuration**
Input	(150, 150, 3)	0	-
VGG16 Base	(4, 4, 512)	14,714,688	**Frozen** (Untrainable)
Flatten	(8192)	0	-
Dense	(1024)	8,389,632	Activation: **ReLU**
Dropout	(1024)	0	Rate: 0.5
Output (Dense)	(1)	1,025	Activation: **Sigmoid**
Total Params		**23,105,345**	**(8.4M Trainable)**

4.4.2 Detection Model: Mask R-CNN Specifications

To provide clinically actionable insights (i.e., localization), an instance segmentation model based on the **Mask R-CNN** framework [34] was implemented. This architecture extends the Faster R-CNN detector by adding a parallel branch for predicting segmentation masks, offering a higher degree of explainability than simple classification. To develop the pneumonia detection model, the best-performing public implementation from the RSNA Pneumonia

Detection Challenge[24] was adapted and extended. The full detection model training notebook and environment is available online[25].

Backbone and Feature Pyramid

The model utilizes a **ResNet-50** [33] backbone combined with a **Feature Pyramid Network (FPN)**. This combination is crucial for medical imaging as the FPN architecture builds high-level semantic feature maps at all scales, enabling the detection of both large consolidations and smaller, subtle focal opacities.

Anchor Box Optimization

Standard anchor box scales were adjusted to align with the typical aspect ratios of lung fields and pneumonia lesions. The Region of Interest (RoI) positive ratio was set to 0.33 to balance the sampling of foreground (pneumonia) and background regions during training. The complete hyperparameter configuration is summarized in **Table 4.2**.

Table 4.2 Model Configuration and Hyperparameters During training

Component	**Configuration**
Framework	Mask R-CNN (Matterport Implementation)
Backbone	ResNet-50 + FPN (Feature Pyramid Network)
Input Policy	Resized to 1024 x 1024
Anchor Scales	16^2, 32^2, 64^2, 128^2 pixels
RoI Positive Ratio	0.33
Classes	2 (Background, Lung Opacity)

[24] https://www.kaggle.com/code/hmendonca/mask-rcnn-and-coco-transfer-learning-lb-0-155
[25] https://www.kaggle.com/code/malaizyogmuzumala/emi-detection-model-mrcnn-and-coco

4.3.3 Multi-Stage Optimization Strategy

Training deep neural networks on specialized medical data requires a careful balance between learning new features and preserving pre-learned knowledge. To achieve this, a **Multi-Stage Optimization Strategy** was employed. This strategy prevents "catastrophic forgetting", where the large gradients from the randomly initialized classifier head destroy the pre-trained weights of the backbone. The training process was divided into three distinct logical stages, as detailed in **Table 4.3**.

Stage I: Network Warm-up (Head Initialization)

In the initial phase, the convolutional backbone (VGG16 or ResNet-50) is 'frozen,' rendering its weights untrainable. Training is restricted exclusively to the newly added classifier and regression heads. A higher learning rate (η = 0.012) is utilized to allow these layers to quickly converge from their random initialization to a stable state.

Stage II: Deep Fine-Tuning

Once the heads have stabilized, the backbone layers are 'unfrozen.' The entire network is then trained end-to-end. Crucially, a significantly lower learning rate ($\eta = 0.006$) is applied to ensure that the pre-trained feature extractors are fine-tuned slowly rather than drastically altered. This stage allows the model to adapt its high-level features (e.g., edge detection) to the specific textures of X-ray radiography.

Stage III: Extended Refinement

The final stage involves extended training with full data augmentation enabled. This phase maximizes the model's generalization capability by exposing the network to the stochastically transformed images generated by the pipeline described in Section 4.3.2.

Table 4.3. Hyperparameter configuration for the detection task at inference

Parameter Category	**Setting / Value**	**Justification**
Input Resolution	256 x 256 pixels	Computational efficiency
Batch Size	8	Single-GPU memory fit
Classes	2	Background + Pneumonia
Max Instances	4	Dataset domain constraint
Detection Confidence	>= 0.78	Minimize false positives
NMS Threshold	0.01	Suppress overlapping boxes

4.5 Application Interface Implementation

While the Deep Learning models provide the diagnostic logic, their clinical utility is contingent upon an effective delivery mechanism. This section details the implementation of the user-facing application, a web-based interface designed to seamlessly integrate AI-assisted interpretation into the radiologist's workstation. The development followed a component-based architecture using the **Next.js**[26] framework, ensuring responsiveness and modularity.

[26] https://nextjs.org/

Table 4.4 Showing Key modifications introduced on recommendation in first focus group

Aspect	**Before**	**After Focus Group 1**	**Rationale for Change**
Findings text description	'This case is suspected positive for pneumonia'	It was changed to "This radiograph is considered suspicious for pneumonia"	Radiologist recommendation to be more inline with 'industry standards.'
Colour of highlighted text	'Positive' was in light red colour	'Considered suspicious' was in yellow	Radiologist recommendation to be more inline with 'industry standards.'
Workflow changed	The 'Check for pneumonia' and 'Show suspicious regions' buttons needed to be clicked	As soon as the DICOM image is uploaded processing automatically begins	Radiologist recommendation to reduce click fatigue.
Dicom Metadata Bar	The scrollable with some metadata hidden	Metadata always visible in the metadata card above results	Radiology resident recommendation for improved visibility of key data to make the interface more useful.

4.5.1 Web Interface Architecture

The client-side application was built to minimize friction in the diagnostic workflow. It features a "single-page application" (SPA) design, where the upload, processing, and results visualization occur within a unified viewport, eliminating the need for page reloads that could disrupt clinical concentration. The full source code and implementation of the web interface is available online[27].

[27] https://github.com/Enterprise-Medical-Imaging-in-Zambia/emiz_project-wp6-demo-frontend

Table 4.5 Showing Key modifications introduced on recommendation in second focus group

Aspect	**Before**	**After Focus Group 2**	**Rationale for Change**
Displayed pathologies	All investigated pathologies were displayed	Only pathologies considered likely are displayed	Expert recommendation to reduce complexity of the interface
Display Order	Pathologies were ordered randomly	Pathologies returned are displayed in order of likelihood	Expert recommendation

Core Components

- **DICOM Ingestion Module:** A dedicated drag-and-drop zone (Figure 4.2) was implemented to handle .dcm files directly. Unlike standard file uploaders, this module parses the DICOM header client-side to extract patient metadata (e.g., Patient ID, Modality) before the image is sent to the server, providing immediate feedback to the user.
- **Results Visualization Layer:** Upon receiving the inference data from the API, this layer dynamically renders the results. For classification tasks, confidence scores are displayed as progress bars. For detection tasks, the backend would send the images with the results burned onto them as the chief means of display, creating an overlay that can be toggled on or off by the clinician to view the results or original image.



Figure 4.2 The file selection module of the web interface, supporting direct DICOM ingestion

4.5.2 User Interface Evolution

The interface underwent significant architectural evolution based on the iterative feedback cycles defined in the methodology. This evolution reflects a shift from a "function-first" prototype to a "workflow-first" clinical tool.

Initial Prototype (v1.0)

The initial release (Figure 4.3) focused on technical verification. It required explicit user interaction for every step: uploading an image, clicking a "Check for Pneumonia" button, and clicking a separate "Show Suspicious Regions" button to visualize localization. While functional, this design introduced unnecessary 'click fatigue.'

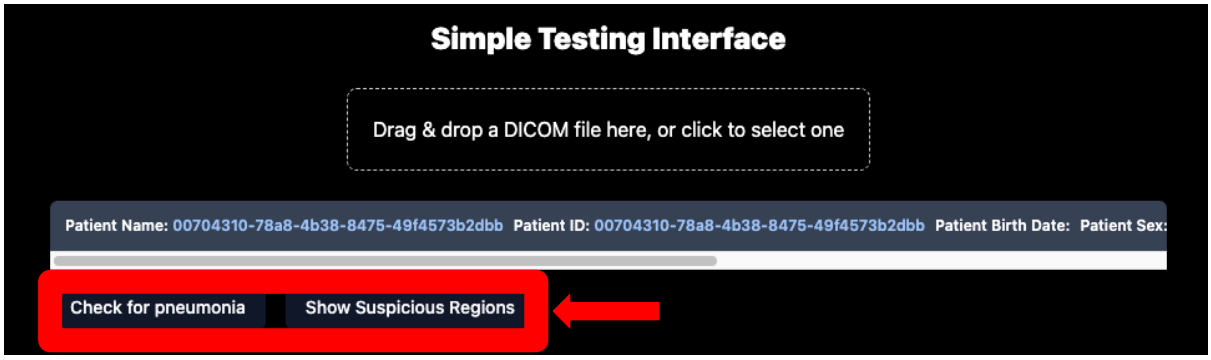


Figure 4.3 Interface prototype requiring manual triggers for inference.

Refined Clinical Interface (v2.0)

Following expert consultation, the interface was refactored (Figure 4.4). The workflow was automated: image upload now triggers an immediate, asynchronous call to the inference API. The visualization logic was updated to use industry-standard colour coding, shifting from a generic red to a 'warning yellow' for suspected regions to align with radiological conventions for cautionary alerts rather than definitive error messages.

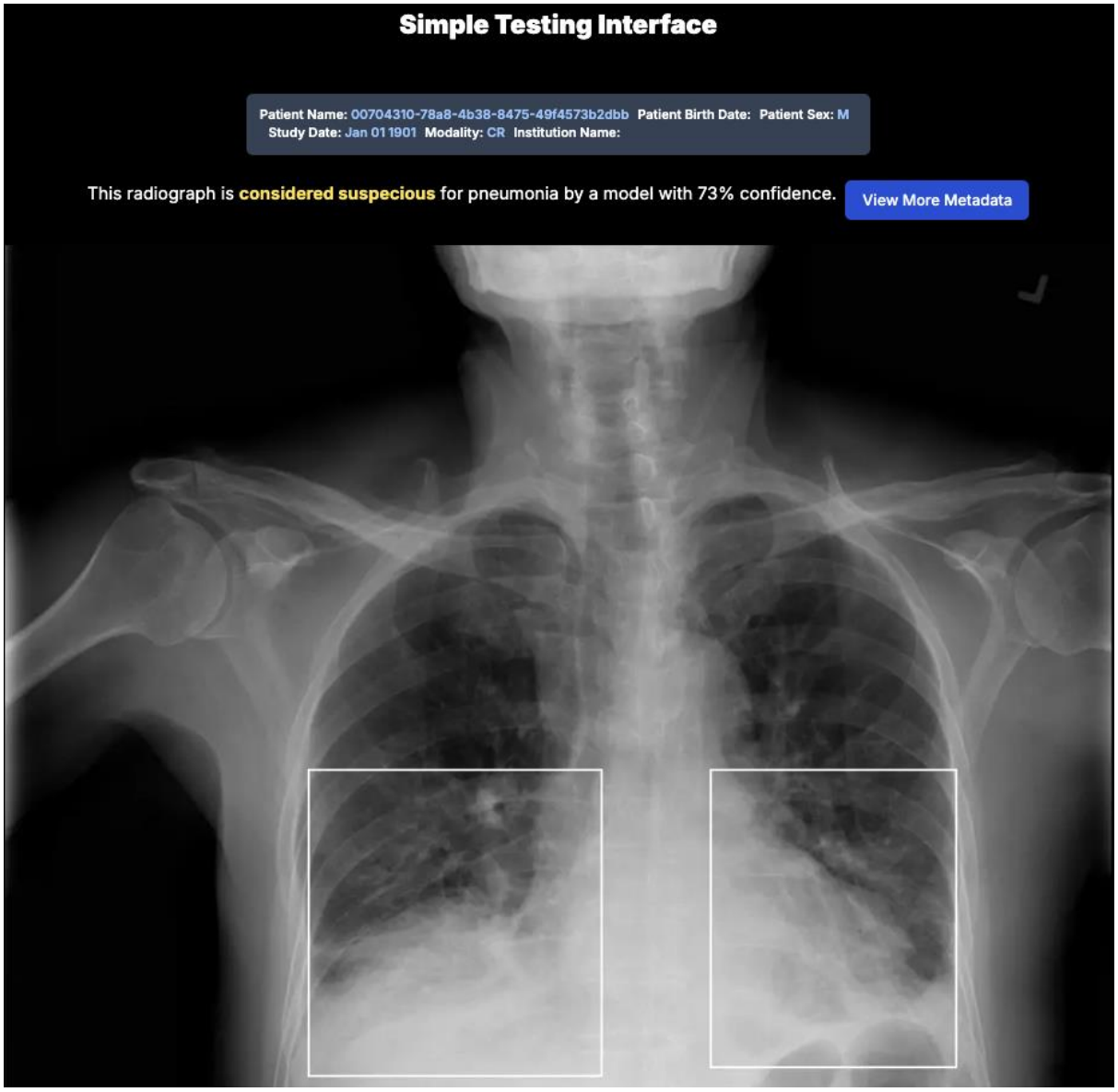


Figure 4.4 The v2.0 interface displaying automated detection results with standardized visualization

Multi-Pathology Dashboard (v3.0)

To address the clinical reality that patients often present with comorbid conditions, the final iteration (Figure 4.5) introduced a multi-pathology panel. This component creates a scalable architecture where the frontend can dynamically render results for multiple disease classes (e.g., Pneumonia, Tuberculosis, Effusion) based on the JSON response array, sorting them by probability to prioritize the clinician's attention.

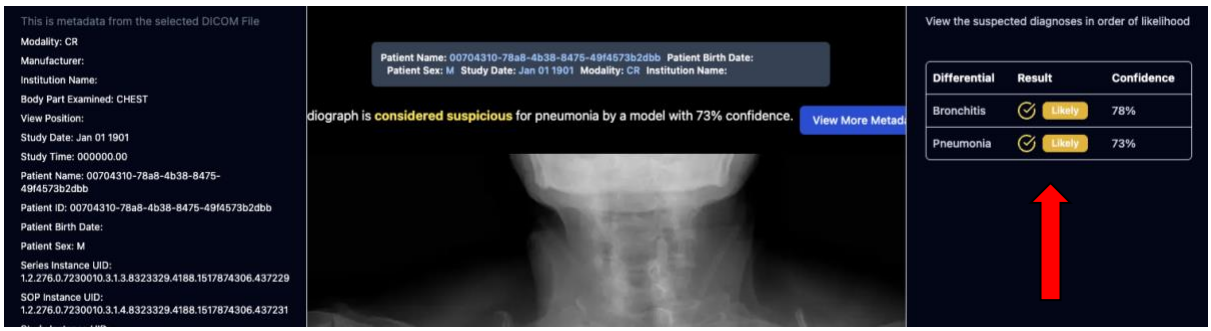


Figure 4.5. The final v3.0 interface features the multi-pathology suspicion panel.

4.6 Server-Side Logic and API Implementation

The backend infrastructure serves as the computational engine of the diagnostic system. Implemented in **Python** using the **Flask** micro-framework, it was designed to be stateless and scalable, capable of handling concurrent inference requests from multiple client interfaces. This section details the API architecture, the endpoint definitions, and the request-response lifecycle. The Flask servers that ran the detection model and that for the classification model ran on the same machine but on different ports. The full source code and implementations of the classification model server[28] and the detection model server[29] are publicly available.

4.6.1 API Endpoint Definition

The server exposes a set of RESTful endpoints, each mapped to a specific diagnostic function. These endpoints utilize standard HTTP methods to ensure compatibility with various client applications, including the web interface and third-party integrations.

The Classification Endpoint

This endpoint facilitates the core classification task. It accepts a POST request containing the binary image file. The server deserializes the image, passes it through the pre-trained VGG16 classifier, and returns a JSON object containing the boolean prediction and the confidence score.

- **Input :** multipart/form-data with the DICOM image (Key: image, Value: image.dcm)

[28] https://github.com/Enterprise-Medical-Imaging-in-Zambia/emiz_project-wp6-binary-classifier-code
[29] https://github.com/Enterprise-Medical-Imaging-in-Zambia/emiz_project-wp6-simple-bounding-box-model

- **Output :** JSON { "pneumonia": true }

The Detection Endpoint

Designed for the detection task, this endpoint handles the more complex localization logic. Upon receiving an image, it invokes the Mask R-CNN model. The response includes not just the classification, but a structured array of coordinates representing the Regions of Interest (RoI).

- **Input:**(Key: file, Value: image.dcm)
- **Output:** JSON payload containing an array of box objects:


```
{
    "predictions": [
        {
            "box": { "x1": 120, "y1": 45, "x2": 380, "y2": 295 },
            "class_id": 1,
            "score": 0.9723
        }
    ]
}
```


4.6.2 Request Validation and System Verification

To ensure the robustness of the API before deployment to the Azure cloud environment, a comprehensive verification suite was conducted using Postman[30], an API development and testing platform.

4.7 Integrations

A critical requirement for the adoption of AI in resource-constrained settings is the ability to integrate new diagnostic tools into existing infrastructure without necessitating a complete overhaul of the clinician's workstation. To demonstrate this interoperability, the core inference API was extended beyond the standalone web interface through two specialized integrations: a plugin for a DICOM server and a browser-based extension.

[30] https://www.postman.com/

4.7.1 Intelligent Orthanc Plugin

Orthanc is a widely adopted, open-source DICOM server used extensively in low-resource settings due to its lightweight architecture and RESTful nature [42]. To bring AI directly into the radiologist's primary environment, a custom plugin was developed.

Implementation Strategy

The integration was achieved by developing a specialized annotation plugin specifically designed for the Orthanc Web Server [91] . This plugin acts as a proxy, intercepting the currently viewed image data and forwarding it to the centralized inference server (described in Section 4.6).

- **Workflow:** The plugin is designed to capture the event of image loading within the viewer. Once triggered, it transmits the medical image data to the inference server either through the plugin itself or via a web service interface, to receive predictions or analysis results.
- **Visualization:** Using the output from the AI model, the plugin dynamically generates annotations in real-time. These findings are displayed as visual indicators (such as bounding boxes) to highlight pathologies directly on the image (Figure 4.8). Additionally, the results are passed as custom metadata tags (e.g., "No Pneumonia") displayed within the viewer's metadata panel. This seamless integration removes the need for physicians to export files and upload them to a separate portal, significantly reducing workflow friction.

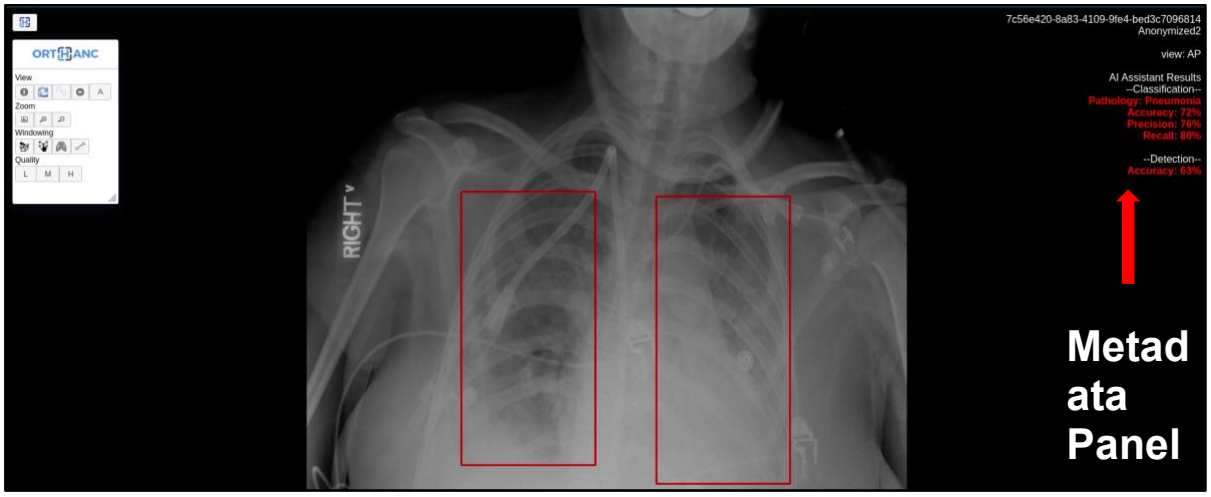


Figure 4.6 The Intelligent Orthanc Viewer plugin displays AI-detected pneumonia regions within the native DICOM viewer interface showing model results in the metadata panel.

4.7.2 Browser Extension

To support generalist doctors who may not have access to a dedicated DICOM viewers workstation but rely on web-based hospital portals, a browser extension was developed for **Google Chrome**[31] and **Mozilla Firefox**[32].

Architecture

The extension was built using the WebExtensions API[33], allowing it to interact with the Document Object Model (DOM) of any active webpage.

- **Context Awareness:** The extension injects a sidebar (Figure 4.7) into the browser window when a medical image format is detected.
- **Functionality:** It provides a lightweight version of the main interface, featuring the metadata extraction panel and the multi-pathology result display. By operating at the browser level, this tool democratizes access to the AI models, allowing clinicians to run an analysis on an X-ray displayed in a basic web-based Electronic Health Record (EHR) system or even a webmail client.

[31] https://www.google.com/chrome/
[32] https://www.firefox.com/en-US/
[33] https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions

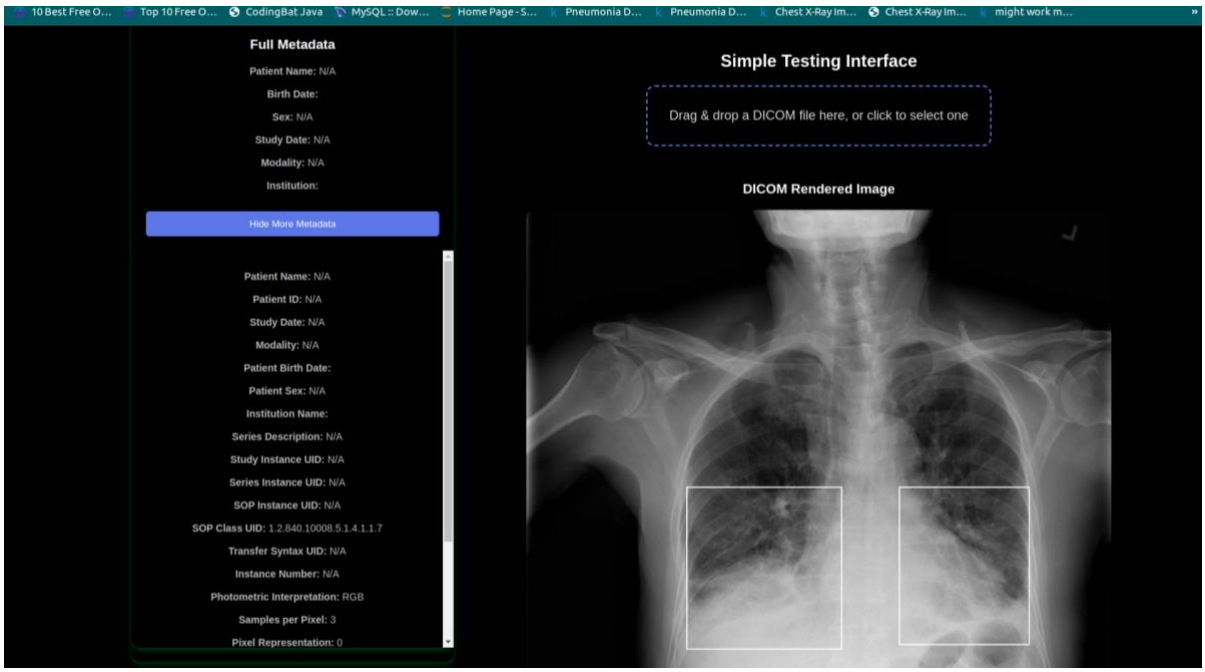


Figure 4.7 The intelligent browser extension analyses a chest radiograph within a Chrome browser session, featuring the metadata side-panel

4.8 Chapter Summary

This chapter detailed the technical development of an AI-assisted radiological interpretation system tailored to the operational realities of Zambian healthcare. Driven by clinician requests for spatial explainability and workflow efficiency, the system utilizes a lightweight, distributed client-server architecture that offloads intensive computation to a centralized cloud instance.

Key technical achievements include:

- **AI Models & Data Processing:** Implemented a VGG16 model for rapid pneumonia classification and a Mask R-CNN model for localized lesion detection. These models were trained using a custom data augmentation heuristic designed to simulate the specific image degradations common in low-resource radiography.
- **Clinician-Centric Interface:** Developed a responsive Next.js web application. Through iterative clinical feedback, the interface evolved into an automated, multi-pathology dashboard that provides ranked diagnostic suggestions and highlights regions of interest.

- **Seamless Integration:** Built a robust Flask-based REST API that powers the main web app alongside two specific integrations: an Orthanc DICOM server plugin and a browser extension. These allow the AI tools to be embedded directly into existing hospital PACS and web portals.

In sum, the chapter demonstrated the successful engineering of a modular, clinically aligned AI tool designed specifically to overcome the technical and workflow constraints of healthcare settings in the Global South.

Chapter 5. Results

This chapter presents the empirical findings derived from the mixed-methods research design outlined in Chapter 3 and the evaluation of the technical artifacts detailed in Chapter 4. While the preceding chapter focused on the engineering implementation and verification of the AI-assisted interpretation system, this chapter focuses on the validation of the research objectives through quantitative statistics and qualitative thematic analysis.

- **Section 5.1: Contextual Inquiry Findings** details the results from the initial clinician survey and semi-structured interviews. This section quantifies the demographic characteristics of the Zambian radiological workforce and statistically analyses existing challenges related to workload, efficiency, and diagnostic effectiveness.
- **Section 5.2: AI Model Performance Metrics** reports the technical validation of the machine learning models. This includes the evaluation of the binary classification and bounding-box detection algorithms against standard performance metrics such as Accuracy, Precision, Recall, and Intersection over Union (IoU).
- **Section 5.3: Experimental Evaluation Results** presents the findings from the controlled user study. It details the impact of the AI intervention on the practitioners' cognitive load, measured via the NASA-TLX, and assesses user acceptance through the TAM 2 constructs.

Collectively, these findings provide the evidence required to address the research questions regarding the feasibility, performance, and acceptance of AI tools in resource-constrained radiological workflows.

5.1 Contextual Inquiry Findings (RQ1)

The initial phase of the research sought to establish a baseline understanding of the current radiological challenges landscape in Zambia (see Section 3.4). This section details the findings from the two studies that were conducted, a set of interviews conducted with medical doctors and radiology residents (see Section 3.4.4) and a survey of clinical practitioners of various types (see Section 3.4.3).

5.1.1 Survey of Clinicians

This section presents the findings from the survey of medical practitioners in Zambia. The survey had 101 responses and simple screening one response did not meet the criteria. It was discarded. The results are organized into four sections: (i) demographic characteristics of respondents, (ii) experience and confidence in medical image interpretation, (iii) perceptions on workload and reporting quality, and (iv) inferential analysis of relationships between key variables. The full demographic details are available in Table 5.1.

5.1.1.1 Demographic Characteristics

A total of 100 responses were analyzed as shown in Table 5.1. Respondents were drawn from facilities across different provinces in Zambia and represented various ownership types and levels (see study setting in Chapter 3.4.3).

Personal Demographics and Education

The gender distribution of respondents was slightly uneven, comprising more males (n=57, 57%) than females (n=43, 43%). The age distribution indicated a predominantly young to mid-career workforce, with the majority (n=82, 82%) aged between 25 and 40 years old. Specifically, the 31–40 age group constituted the majority of the sample (n=52, 52%), while the 25–30 group made up the next largest group (n=30, 30%). Regarding educational attainment, the majority of participants (n=67, 67%) held a bachelor's degree, followed by 21 percent with a master's degree. Only one percent held a PhD, with the remainder holding diplomas or advanced diplomas. Regarding years of practice, the largest cohort had between two and five years of experience (n=40, 40%). This was followed by participants with 11 - 20 years (n=19, 19%), 6–10 years (n=17, 17%), and less than one year (n=15, 15%). The smallest group had more than 20 years of experience (n=4, 4%).

Facility Characteristics

The vast majority of respondents were in GRZ facilities (n=75, 75%), reflecting the public sector's dominance in Zambian healthcare provision. The remaining participants worked in Private (n=9, 9%), Military (n=9, 9%), and NGO (n=7, 7%) facilities. Geographically, there was a significant concentration of respondents in Lusaka Province (n=60, 60%), consistent with the centralization of tertiary and specialized healthcare services in the capital, followed

by the Copperbelt (n=18, 18%). Regarding current facility levels, the majority of participants were from hospitals (n=77, 77%) with level 3 hospitals forming the largest group of those (n=37, 37%), there was lower representation from other facilities with Diagnostic Centre being the lowest represented (n=3, 3%) of the facility types set as options (see Chapter 3.4.2).

Professional Distribution and Workforce Shortage

A critical finding regarding the 'workforce shortage' is evident in the distribution of professions. The sample was heavily dominated by non-specialist **Medical Doctors** (all specialties excluding radiology), who accounted for the majority (n=65, 65%) of the respondents. In stark contrast, specialized radiological expertise was minimal: **Radiology Residents** accounted for a fraction of the respondents (n=7, **7%)**, and only one participant (n=1,1%) identified as fully qualified **Radiologists**. **Medical Imagers** (including radiographers and technicians) comprised **15%** of the sample.

This disparity, where generalist doctors outnumber qualified radiologists by a ratio of nearly 65:1 within the sample, underscores the severe scarcity of specialized personnel. It highlights a healthcare context where the burden of medical image interpretation disproportionately falls upon general practitioners and residents rather than dedicated specialists.

Table 5.1. Demographics of Clinicians Survey Participants

Variable	Category	Frequency (n)	Percentage (%)
Gender	Male	57	57%
	Female	43	43%
Age Group	18–24	2	2%
	25-30	30	30%
	31-40	52	52%
	41-50	14	14%

	51–60	1	1%
	60+	1	1%
Current Facility Level	Diagnostic Centre	3	3%
	General Clinic	5	5%
	Health Centre	4	4%
	Level 1 Hospital	24	24%
	Level 2 Hospital	16	16%
	Level 3 Hospital	37	37%
	Other	11	11%
Facility Province	Lusaka	60	60%
	Central	4	4%
	Copperbelt	18	18%
	Eastern	2	2%
	North-Western	9	9%
	Southern	6	6%
	Western	1	1%
Participant Facility Ownership	GRZ	75	75%

	Private	9	9%
	NGO	7	7%
	Military	9	9%
Highest Level of Education	Advanced Diploma	2	2%
	Bachelors Degree	67	67%
	Diploma	9	9%
	Master's Degree	21	21%
	PhD	1	1%
Profession	Clinical Officer/Medical Licentiate	9	9%
	Medical Doctors (All specialties aside radiology)	65	65%
	Medical Imager (Radiographer, Radiography Technician, Sonographer, Radiation Physicist, etc)	15	15%
	Medical Officer	1	1%
	Nurse	1	1%
	Radiologist	1	1%
	Radiology Resident	7	7%
	Strategic Information Department	1	1%

Years of Practice	<1	15	15%
	2–5	40	40%
	6-10	17	17%
	11-20	19	19%
	> 20	4	4%

5.1.1.2 Clinical Experience and Confidence Levels

The survey results reveal a healthcare landscape where medical image interpretation is a ubiquitous task, performed by the vast majority of practitioners regardless of their specialization. This finding confirms that in the absence of radiologists, the responsibility for radiological diagnosis has shifted almost entirely to the general workforce.

Modality Exposure and Complexity

Experience levels varied significantly by imaging modality. Plain X-ray radiographs were the most universally encountered modality, with 96.8 percent (n=91) of experienced respondents reporting engagement with them. Exposure to more complex modalities dropped notably, with Ultrasound at 59.6 percent (n=56), Computed Tomography (CT) at 47.9 percent (n=45), and Magnetic Resonance Imaging (MRI) at only 17.0 percent (n=16).

This distribution highlights a critical operational challenge: while nearly all doctors interpret X-rays, nearly half of the workforce is also attempting to interpret complex CT scans. This occurs in a system where only seven percent of the sample are radiology residents and one percent are radiologists, suggesting that complex cross-sectional imaging is frequently interpreted by generalists without specialized training.

The Confidence-Competence Mismatch

A critical analysis was conducted to examine the relationship between modality experience and self-reported confidence. While general confidence levels were moderate (Mean = 3.27, SD = 0.69), a deeper cross-tabulation of *Experience* versus *Confidence* reveals a concerning "confidence-competence mismatch," among general practitioners handling complex imaging.

Specifically, while nearly all doctors interpret X-rays, a significant subset of Medical Doctors (n=30) reported having experience interpreting complex modalities (CT and MRI). Crucially, 73% of this specific group rated their confidence as only "average" or "low" (Likert scale ≤ 3).

This disparity indicates that the specific cadre of clinicians most heavily relied upon to fill the workforce gap is frequently required to interpret high-complexity studies despite possessing low diagnostic confidence. As shown in Table 5.2, the high experience rates in complex modalities do not correlate with high confidence scores among generalists.

Table 5.2 Cross-tabulation of Experience with Complex Modalities (CT/MRI) versus Self-Reported Confidence Levels among Medical Practitioners.

		Reported Level of Confidence		
		Participants Interpreting Complex Modalities (CT/MRI)	Low / Average (Likert Scale 1 - 3)	High / Very High (Likert Scale 4–5)
Profession	General medical Doctors	n=30	73.3%	26.7%
	Radiologists and residents	n=8	50%	50%

Implications for AI Intervention

This finding presents the strongest argument for the introduction of AI decision support systems. While a significant percentage of general doctors are operationally required to interpret complex scans due to the shortage of specialists, the majority do not feel highly confident in doing so. This mismatch highlights a critical safety gap in the current workflow:

complex diagnostic decisions are being made with suboptimal confidence, creating a high-risk environment that AI tools could effectively bridge by providing second opinions and diagnostic validation.

5.1.1.3 Perceptions of Workload and Time Pressure

The survey data paints a picture of a healthcare system where study volumes are directly compromising diagnostic quality. As illustrated in Figure 5.1, there is a strong consensus regarding the negative impact of workload on output. A substantial majority of respondents (n=64, 68.1%) agreed or strongly agreed that the volume of studies they interpret negatively affects the quality of their reports. This suggests that practitioners are acutely aware that the current demand exceeds their capacity to maintain high clinical standards.

As shown in Figure 5.1, a majority of respondents agreed that high study volumes affect the quality of reports, with 68.1 percent agreeing or strongly agreeing. Surprisingly, 31.0 percent indicated that they often have more imaging studies to interpret than can reasonably be handled in a workday which was a minority. A majority of all respondents (74.5%) agreed that their effectiveness is reduced when patient clinical history is incomplete or missing. Confidence in consistently accurate interpretation had a majority having positive responses, with 58.5 percent reporting agreement, while 16.0 percent indicated disagreement.

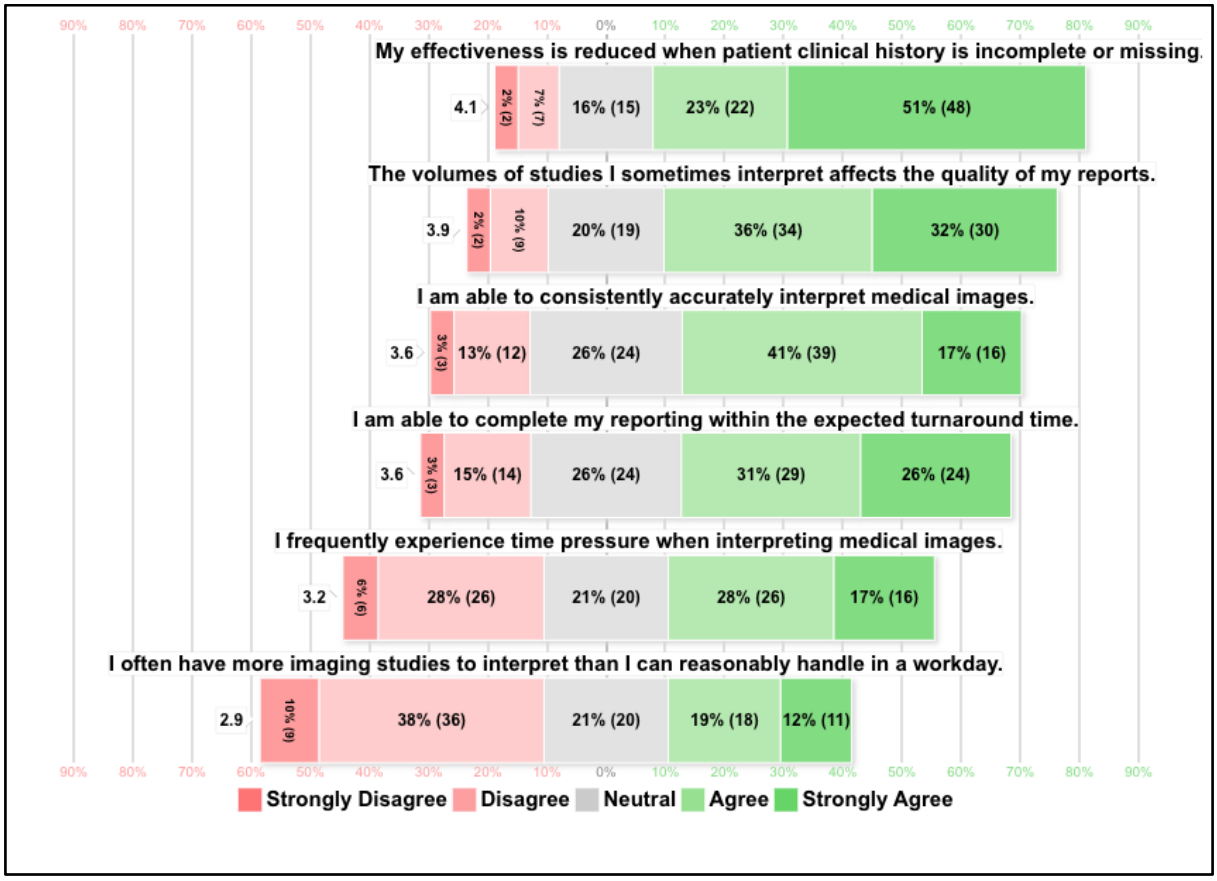


Figure 5.1. Charts of Likert responses to workload and effectiveness statements

To facilitate analysis, a workload index was created as described in chapter 3.8.4.

The Burden on Specialists versus Generalists

While the aggregate data suggests a moderate workload burden across the board, a deeper stratification by profession reveals a stark structural inequality in how this pressure is distributed. The survey data identifies a clear 'imaging bottleneck': the professionals most qualified to interpret images are the ones most overwhelmed by the volume.

Descriptive analysis showed that **Radiology Residents and Radiologists** reported the highest Workload Index mean (**4.08**), closely followed by **Medical Imagers** (**3.87**). In contrast, **General Medical Doctors** reported a significantly lower mean (**3.12**). This disparity was confirmed to be statistically significant via the Kruskal-Wallis test ($H(3) = 12.42$, $p = .006$).

To isolate exactly where these differences lay, Dunn's post-hoc pairwise comparisons were conducted (see **Table 5.**3). The results indicate that the perceived workload for professionals dedicated to imaging (Residents and Imagers) is significantly higher than that of general

practitioners. The effect sizes (r_{rb}) for these comparisons were **0.52** and **0.47** respectively, which are considered moderate-to-large effects.

This finding tells a critical story about the Zambian radiological workflow: General doctors, who likely interpret images sporadically as part of broader patient management, feel less pressure regarding the quantity of images. However, the specialized workforce (which serves as the referral endpoint for difficult cases) is operating under severe strain. This suggests that the 'shortage of radiologists' is not just a headcount issue, but an operational reality where the few available experts are saturated, validating the need for AI tools that can triage routine cases to relieve this specific bottleneck.

Table 5.3 Pairwise Comparisons of Perceived Workload between Professional Groups (Dunn's Post-Hoc Test).

Comparison Pair	**Z-Statistic**	**p-value (p_{holmp})**	**Effect Size (r_{rb})**	**Interpretation**
Radiology Experts vs. Medical Doctors	-2.578	.050	0.52	Significant: Experts perceive substantially higher workload.
Medical Imagers vs. Medical Doctors	-2.681	.044	0.47	Significant: Imagers perceive substantially higher workload.
Medical Imagers vs. Radiology Experts	-0.335	1.000	0.24	Not Significant: Both specialized groups face similarly high pressure.

Systemic Nature of the Challenge

Notably, this perception of workload burden appears to be a systemic issue inherent to the Zambian radiological workflow, rather than a transient challenge faced only by junior staff. Statistical analysis of **Years of Practice** revealed no significant difference in workload perception across experience levels ($H = 3.33$, $p = .504$)

An analysis of the descriptive statistics shows, practitioners with two to five years of experience reported a Workload Index ($M = 3.28$, $SD = 0.93$) nearly identical to one with 11 - 20 years of experience ($M = 3.23$, $SD=0.93$). This uniformity suggests that the inefficiencies identified in the qualitative phase (such as the reliance on physical media and manual reporting processes) exert a constant pressure that clinical experience alone cannot mitigate. The burden is structural; senior clinicians are not insulated from the time pressures of a manual workflow. This finding critically validates the research premise: the solution to these bottlenecks lies not merely in personnel training, but in the introduction of efficiency-enhancing tools like AI and PACS to address the systemic barriers affecting the entire workforce.

5.1.1.4 Challenges to Effectiveness and Efficiency

Beyond the sheer volume of work, the survey results highlight distinct barriers that degrade the effectiveness of the interpretation process itself. The data points to a workflow that is not only busy but also operationally fragmented.

The Systemic Barrier of Missing Clinical History

The most pronounced challenge to diagnostic effectiveness identified in the survey was the lack of contextual patient information. As shown in Figure 5.1, **74.5 percent** of respondents agreed or strongly agreed that their effectiveness is reduced when patient clinical history is incomplete or missing. With a mean agreement score of **4.14 (SD = 1.07)**—the highest among all effectiveness items—this finding points to a critical breakage in the diagnostic workflow. Without the requisite clinical context, practitioners are often forced to interpret imaging findings in a vacuum, significantly increasing the cognitive load and the risk of diagnostic error.

Crucially, statistical analysis reveals that this challenge is uniform across the entire medical workforce. A Kruskal-Wallis test examining **Years of Practice** found no statistically

significant difference in how this barrier is perceived (H = 3.387, p = .495). Similarly, no significant differences were found across **Gender** (p = .756). This lack of variance is a key finding: it indicates that the absence of clinical history is a **structural and systemic flaw** in the Zambian radiological workflow, rather than an issue related to individual practitioner experience or competence. Whether a respondent was a junior house officer or a senior consultant, the operational failure to convey patient history remains a universal impediment to effective diagnosis.

Efficiency and Turnaround Time

While a slight majority of all respondents (56.4%, n=53) agreed that they are able to complete their reporting within expected turnaround times, a deeper stratification of the data reveals that this efficiency is not evenly distributed across the workflow. The statistical analysis of this metric (Kruskal-Wallis H = 11.99, p = .007) uncovered a distinct operational divide between image acquisition and image interpretation.

Post-hoc analysis with Holm correction indicated that **Medical Imagers** (Radiographers/Technicians) reported significantly higher agreement (M=4.39, SD=0.87) regarding their ability to meet turnaround times compared to **Medical Doctors** (M=3.40, SD=1.152), (p_{holm} = .018). This disparity represents a large effect size (r_{rb} = .498) and highlights a critical workflow dynamic: the bottleneck in the Zambian radiological system lies not in the acquisition of images, but in their interpretation.

Furthermore, **Radiology Experts** reported the lowest mean agreement (M=3.25) of all groups. This finding suggests that the specialized workforce tasked with the most complex reporting is operating under the greatest time pressure. The data confirms that while the technical process of capturing images is relatively efficient, the cognitive task of diagnostic reporting is the primary constraint slowing down the system, reinforcing the need for AI solutions specifically targeted at accelerating the interpretation phase.

Demographic Uniformity of Challenges

Critically, the perception of effectiveness barriers was found to be uniform across the workforce demographics. Inferential statistical tests applied to the items regarding **'Missing Clinical History'** and **'Consistent Accuracy'** revealed no significant differences based on **Years of Practice** ($p > .05$) or **Gender** ($p > .05$).

For example, the sentiment that missing history reduces effectiveness was shared equally by junior doctors with less than one year of experience ($M=4.21$, $SD=1.25$) and veteran practitioners with over twenty years of experience ($M=4.25$, $SD=0.957$). This lack of variance is a significant finding in itself: it demonstrates that the barriers to effectiveness are **structural flaws inherent to the healthcare system**, rather than challenges that practitioners can 'outgrow' with experience. The universality of these issues across age, gender, and experience levels provides robust evidence that individual training interventions alone are insufficient. Instead, the data validates the necessity for systemic, technological interventions—such as AI-driven decision support—that address the root causes of workflow inefficiency for all cadres of staff.

5.1.2 Interviews with Medical Doctors and Open Ended Survey Responses

As outlined in Section 3.4.4 interviews were conducted with three general practitioner medical doctors and one radiology resident. Subsequently analysis was conducted as outlined by Braun and Clarke [11], see Chapter 3.4.5. The findings are categorized into four themes, each explicitly mapping to the core dimensions of the Research Questions (RQ). The first three themes are concerned with workflow *efficiency*, diagnostic *effectiveness*, and cognitive *workload*. These themes illustrate how infrastructural and systemic deficits create a cycle of inefficiency that ultimately compromises patient care and addresses RQ1. The fourth theme around criteria for interventions and requirements and limitations addresses RQ2 and RQ3.

Table 5.4 Interview and Survey Open Response Participant Profile Table

Role	Station	Years of Practice	Participant Code
Radiology Resident	UTHs Adult Hospital	5	RaRe
Senior Resident Medical Officer	UTHs Children's Hospital	3	SRMO
Senior Health Officer	UTHs Adult Hospital	4	SeHO
Junior Resident Medical Officer	UTHs Adult Hospital	2	JRMO1
Clinical Officer / Medical Licentiate	Maina Soko Medical Center	11-20	COML1
Medical Doctor	Kitwe Teaching Hospital	< 1	MD1
Medical Doctor	Chelstone Zonal Hospital	6 - 10	MD2
Medical Doctor	UTHs Adult Hospital	2 - 5	MD3
Medical Imager	Special Emergency Services	2 - 5	MI1
Consultant Doctor	UTHs	11 - 20	CD1
Medical Doctor	UTHs Adult Hospital	6 - 10	MD4

5.1.2.1 Theme 1: Infrastructure-Driven Inefficiency: Suboptimal Access and Management of Digital Images

This theme addresses the research question regarding **efficiency** and **effectiveness** by highlighting how physical infrastructure constraints degrade the speed and quality of image interpretation. Despite the digital nature of modern modalities like CT and Digital Radiography, the workflow remains heavily reliant on analogue or physical transfer mechanisms, creating significant bottlenecks.

The 'Analog' Transfer of Digital Data

A primary barrier to **efficiency** is the reliance on physical media (Compact Discs) and manual transport for data access. The current workflow does not leverage networked PACS. Instead,

as [RaRe] described, the process involves a manual chain of custody: ‘We have patients that would walk into the department... an image would be taken... then from that point, the patients will be presented with a CD, and that CD together with the request form is going to find its way on my desk.’

This physical transfer creates a single point of failure. If a disc is damaged, lost, or incompatible with the viewing computer, the diagnostic data is effectively lost, requiring re-imaging or causing indeterminate delays. The lack of a seamless digital network means that "time is lost" in the burning and loading of images, directly contrasting with the urgent clinical need for rapid diagnosis.

Degradation of Diagnostic Effectiveness via Non-Optimized Hardware

The interview data revealed a critical threat to **effectiveness** regarding the hardware used for interpretation. In the absence of medical-grade monitors, practitioners frequently resort to ‘secondary capture’, viewing images on non-dedicated devices such as mobile phones or capturing photographs of a screen. [SRMO] provided a vivid account of how this practice introduces artifacts that mimic pathology, stating: ‘You find that those... lines that pop up... And some distortion in... contours… light and dark areas mean something, but if you capture it in such a way that it creates a fake light or darkness on it, you may read it wrongly.’ Similarly, [SeHO] noted the risk of looking at ‘an image of an image,’ which inherently reduces resolution and contrast resolution. This finding helps explain the "low confidence" reported in the survey; practitioners are forced to make high-stakes decisions based on visually compromised data.

5.1.2.2 Theme 2: Workflow Fragmentation: Barriers to Effectiveness and Critical Delays

This theme directly addresses the challenges of diagnostic **effectiveness** caused by the decoupling of clinical context from imaging data, and the loss of **efficiency** due to delays in specialist reporting. The qualitative data reinforces the survey finding that missing clinical history is a pervasive systemic failure.

The Diagnostic Void of Missing History

Effectiveness is severely compromised by the fragmentation of information. The ‘analogue’ nature of the request process, where hand-written forms are physically carried by patients, often

results in incomplete data transfer. A [RaRe] participant emphasized that without this context, distinguishing between similar pathologies becomes impossible: ‘If the clinical history is missing, we have a lot of mimickers on imaging... to differentiate the two, is that we have to marry the clinical history.’ This lack of integration forces the radiologist to operate in a diagnostic vacuum, reducing the report's precision and potentially leading to ambiguous or incorrect conclusions.

Operational Delays and Patient Outcomes

The severe shortage of radiologists creates a bottleneck that drastically reduces the system's **efficiency** for complex cases. Routine interpretation is shifted to generalists, but for high-stakes diagnostics like cancer staging or CT scans, the wait for a specialist report can be critical. [SRMO] reported turnaround times of ‘three, four, five days’ for CTs and up to ‘two weeks’ for cancer staging reports.

The impact of this inefficiency extends beyond mere inconvenience; it has direct clinical consequences. As the [RaRe] summarized with the adage ‘Time is brain,’ the inability to secure a timely report leads to delayed interventions. The [SRMO] corroborated this, noting that such ‘delays treatment because you're not sure what you're dealing with,’ forcing clinicians to manage patients empirically while awaiting diagnostic confirmation. This theme underscores that the ‘shortage of radiologists’ is not just a staffing metric but a direct driver of poor patient outcomes.

5.1.2.3 Theme 3: Cognitive Overload: The Burden of Research and Consultative Support

This theme illuminates the **workload** implications of the current system. It reveals that ‘workload’ is not merely a function of volume (number of images) but of the high cognitive effort required to reach a confident diagnosis in a low-resource setting without integrated decision support.

The Hidden Workload of ‘Research’

Practitioners described a workflow where interpretation is not a solitary event but a research-intensive process. A [RaRe] participant noted that research is routine: ‘We still get to research on different sites... Google Scholar, Radiopedia.’ This constant need to verify findings against

external literature adds significant time to each case, increasing the cognitive load per exam. For generalists like the [SRMO], this creates a conflict between the need for accuracy and the pressure of volume; while acknowledging the necessity of research, he admitted, 'I'd be lying if I say I spend so much time... sometimes you push it aside,' illustrating how excessive cognitive workload can force compromises in diagnostic rigor.

Reliance on Ad-Hoc Peer Review

In the absence of automated decision support or readily available senior radiologists, non-specialists rely on informal, peer-based safety nets to manage this workload. The [SeHO] described a self-imposed 'rule of two,' mandating that a second viewer always check the image. While clinically prudent, this strategy effectively doubles the manpower required for a single diagnosis, further straining the limited workforce. The [SRMO] highlighted that diagnostic accuracy is heavily limited by 'individual exposure,' meaning a lack of prior experience with a specific pathology leads to uncertainty. This reliance on personal experience over systematic support highlights the intense cognitive pressure on junior doctors, who must act as the primary interpreters despite their self-identified lack of confidence.

5.1.2.4 Theme 4: Criteria for Intervention: Requirements and Limitations of Digital and Automated Tools

This final theme directly addresses RQ2 and RQ3 by identifying the specific functional and non-functional criteria required for an AI solution to be effective and adaptable within the Zambian context. The interview data shifts the focus from current challenges to future requirements, outlining a user-driven specification for AI tools.

Technological Friction and Resource Constraints

The participants emphasized that any proposed AI solution must account for the harsh reality of the current technological infrastructure. The [RaRe] participant described existing hospital computers as 'very slow,' with exams sometimes taking 'an hour or two' to load. Furthermore, the sustainability of software is a major concern; proprietary tools often fail because 'the subscription is going to run out,' forcing doctors to pay out-of-pocket or lose access. This establishes a clear criterion for adaptability: effective AI tools in this context must be

computationally lightweight to run on older hardware and, crucially, should minimize reliance on expensive, recurring licensing models (favoring open-source or perpetual licensing).

The Need for 'Augmented' Intelligence, Not Replacement

Participants rejected the notion of 'black box' automation in favour of tools that function as cognitive aids. The [SRMO] conceptualized the ideal AI not as a final decision-maker, but as a 'spark plug' or 'hint' to initiate the diagnostic process. Specific requests for automation targeted high-volume, high-urgency pathologies such as 'pneumothorax,' 'fractures,' and 'pneumonias.' This aligns with the need to reduce **workload** by automating routine detection tasks, allowing the clinician to focus on complex synthesis.

Demand for Transparency and Contextual Relevance

A critical criterion for selection (RQ 3) is the explainability and local validity of the AI. Physicians expressed scepticism about tools trained solely on foreign populations. The [RaRe] participant argued that a tool trained on 'Western countries' data might fail to extrapolate to 'the kind of conditions that we have in this space,' such as the high prevalence of Tuberculosis and its specific radiographic presentations. Moreover, there is a strong demand for explainability; the AI must justify its output. As the [SRMO] demanded: 'It shouldn't just say, oh, this is TB. No, it should say because of that line... that whitish thing.' This highlights that for an AI tool to be accepted, it must be transparent in its reasoning and validated against local epidemiological data.

5.1.2.5 Themes from Open-Ended Survey Comments

The open-ended comments from the clinician survey further illuminated the participants' attitudes toward AI integration, generally coalescing into three distinct perspectives: **Optimistic Pragmatism**, **Evidence-Based Skepticism**, and **Ethical Accountability**.

Many respondents expressed a view of **Optimistic Pragmatism**, recognizing that AI lacks the capacity to conduct physical examinations or gather holistic histories, yet as [COML] stated, viewing the combination of 'professional skills and AI' as a pathway to improved service delivery. This group anticipates AI becoming, as [MI1] puts it, a 'very good aid' for radiologists to produce more concise and effective reports.

Conversely, a theme of **Evidence-Based Scepticism** emerged, with [MD2] explicitly requesting studies ‘proving or disproving AI interpretation’ and expressing uncertainty about clinical accuracy. A participant notably reported a failed attempt to use a general-purpose LLM (ChatGPT) for interpretation, highlighting a gap between interest and the availability of specialized tools. Participant [MD1] stated ‘it didn't even attempt citing it wasn't permitted to.’

Finally, concerns regarding **Ethical Accountability** were stark, with [CD1] asserting that AI can ‘never be relied on’ fully because it cannot be held ethically liable for misinterpretation. Collectively, these comments reinforce the need for AI tools that are not only technically accurate but also transparent, clinically validated, and ethically framed as supportive aids rather than autonomous agents.

5.1.2.6 Summary of Contextual Inquiry Findings

In conclusion, the findings from both the survey and interviews converge to answer **Research Question One**: The challenges in Zambian radiological workflows are characterized by a severe **workload** imbalance, where generalists interpret complex images with low confidence; operational **inefficiency** driven by analogue data transfer and delayed specialist reporting; and compromised **effectiveness** due to missing clinical history and suboptimal viewing hardware.

Regarding **Research Question two** and **three**, the data indicates that effective AI solutions must be designed as ‘Augmented Intelligence’ systems, acting as a ‘second pair of eyes’ or a triage mechanism rather than an autonomous diagnostician. The specific criteria for selection include **computational efficiency** (to run on existing hardware), **explainability** (to justify findings to sceptical users), **local context awareness** (training on local pathologies), and **ethical clarity** (maintaining the physician's accountability). These findings directly informed the design of the AI models and interface detailed in the subsequent implementation chapters.

5.2 AI Model Performance Metrics (RQ2)

This section presents the technical evaluation of the developed AI models, addressing the requirement for locally adaptable and effective solutions identified in the contextual inquiry. To validate the "effectiveness" component of **Research Question 2**, the performance of both the **Pneumonia Classification (VGG16)** and **Localization (Mask R-CNN)** models was

rigorously assessed against standard machine learning metrics. These results determine the baseline reliability of the tools prior to their deployment in the user-centred experimental study. The metrics reported here, Accuracy, Precision, Recall, and IoU, provide the quantitative benchmark necessary to assess whether the models meet the "spark plug" or triage criteria requested by clinicians.

5.2.1 Classification Model Performance Results

The binary classification model, based on the VGG16 architecture, was evaluated on the test subset of the Paul Mooney Chest X-ray dataset. The model achieved an overall **Accuracy of 72 percent** in distinguishing between normal and pneumonia-affected radiographs. While this accuracy indicates a functional baseline for screening, the more granular metrics reveal specific strengths and limitations relevant to clinical deployment.

As detailed in **Table 5.5**, the model demonstrated a **Precision of 76 percent** and a **Recall of 80 percent**, resulting in an **F1-Score of 78 percent**. The relatively higher recall (sensitivity) is particularly significant in a screening or triage context, where the clinical priority, as emphasized by the interview participants, is to minimize false negatives (‘missing a case’). A recall of 80 percent suggests that the model is reasonably effective at flagging potential positive cases for review, aligning with the user requirement for a ‘hint’ or ‘second look’ tool. However, the accuracy level also highlights the necessity for radiologist oversight, reinforcing the ‘Human-in-the-loop’ model proposed in this study rather than fully autonomous diagnosis.

Table 5.5 Performance Metrics for the Pneumonia Classification Model (VGG16).

Metric	Score
Accuracy	72%
Balanced Accuracy	66%
Precision	76%
Recall	80%
F1-Score	78%

5.2.2. Detection Model Performance Results

The localization performance of the Mask R-CNN model was evaluated to determine its efficacy in highlighting specific regions of interest within the chest radiograph. This capability directly addresses the qualitative request from clinicians for a tool that can 'highlight the areas I'm supposed to pay particular attention to.'

The model's performance was assessed using **IoU**, **Precision**, and **Recall**, utilizing a 30 percent overlap threshold to define a correct detection. As presented in **Table 5.6**, the model achieved a **Recall of 79 percent**, indicating a strong capacity to identify relevant pathological features when they are present.

However, the model demonstrated a low **Precision (15%)** and **IoU (12%)**. In the context of computer vision, these figures indicate a tendency towards 'over-detection', generating bounding boxes that may be larger than the lesion or flagging regions that are ultimately benign (false positives). While a precision of 15 percent would be suboptimal for an autonomous system, in a human-in-the-loop workflow, this trade-off favors patient safety. The high recall ensures that potential pathologies are rarely missed, effectively serving the 'spark plug' function requested by users. The lower precision implies that while the model successfully draws attention to the pathology, it requires the clinician's expertise to filter the false positives and refine the exact localization.

Table 5.6 Performance Metrics for the Pneumonia Detection Model (Mask R-CNN).

Metric	Score
Recall	79%
Precision	15%
Intersection Over Union	12%

5.3 Experimental Evaluation Results (RQ3)

This section details the results of the quantitative controlled experiment designed to validate the practical efficacy and user acceptance of the AI-assisted interpretation interface. Moving beyond the algorithmic performance metrics reported in Section 5.2, this evaluation focuses on the HCI dimension, specifically, whether the introduction of the AI tool reduces the cognitive burden on practitioners and meets the usability standards required for clinical adoption in Zambia.

As described in Chapter 3.8.1, to isolate the impact of the AI intervention, a counterbalanced, within-subjects (repeated measures) design was employed. This allowed for a direct comparison of the same practitioner's experience when interpreting medical images under two distinct conditions: a standard manual workflow and an AI-assisted workflow. The interpretation task involved reviewing chest radiographs suspected of pneumonia to determine both the presence of pathology and the anatomical regions involved.

The findings are presented in the following sequence:

1. **Participant Demographics (5.3.1):** A profile of the medical practitioners involved in the study.
2. **Impact on Perceived Workload (5.3.2):** An analysis of cognitive load reduction based on the NASA Task Load Index (NASA-TLX).
3. **User Acceptance and Usability (5.3.3):** An evaluation of Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) using the Technology Acceptance Model 2 (TAM 2).
4. **Qualitative Feedback and Feature Requests (5.3.4):** A thematic analysis of open-ended practitioner feedback regarding the tool's clinical utility and limitations.

5.3.1 Participant Demographics

A total of 47 medical practitioners participated in the controlled evaluation study. The sample was purposively selected to represent the operational workforce involved in medical image interpretation within Zambian public hospitals.

Consistent with the analysis strategy defined in Section 3.8.4, the cohort was stratified into two primary groups to facilitate robust statistical comparison:

1. **Specialists (n=16):** Comprising the qualified Radiologist and Radiology Residents.
2. **Generalists (n=31):** Comprising Medical Doctors, Senior Residents, and Health Officers from non-radiological specialties.

The demographic distribution is summarized in Table 5.7. Notably, the majority of participants (81%) possessed between one and five years of clinical experience. This indicates a predominantly early-career workforce, a demographic highly relevant to the study as they are the primary targets for AI-based decision support scaffolding.

Table 5.7. Demographics of Participants in the AI Tool Evaluation Experiment

Variable	**Category**	**Frequency (n)**	**Percentage (%)**
Gender	Male	29	62
	Female	18	38
Profession	Radiologist	1	2
	Radiology Resident	15	32
	Medical Doctor (Other specialties)	31	66
Years of Practice	< 1	1	2
	1 - 5	38	81
	5 - 10	8	17

5.3.2 Results of workload assessment

To evaluate the 'cost' of using the AI tool—specifically the cognitive and physical burden placed on the clinician—the NASA Task Load Index (NASA-TLX) scores were compared between the manual and AI-assisted workflows.

Preliminary testing for normality on the difference scores (NASA_DIFF) using the Shapiro-Wilk test yielded a significant result ($W = 0.918$, $p = .017$), indicating a non-normal distribution. Consequently, the non-parametric **Wilcoxon Signed-Rank Test** was employed to analyse the paired workload data.

5.3.2.1 Overall Workload Assessment

Descriptive statistics indicated a reduction in the global workload when using the AI tool. The mean unweighted Global Workload Score for the AI-assisted condition was **1.74** (SD = 1.03) compared to **2.08** (SD = 1.08) for the standard manual workflow.

However, inferential analysis revealed that while there was a directional trend toward reduced workload, this difference did not reach the standard threshold for statistical significance (W = 295.0, z = 1.676, p = .096). This suggests that while practitioners generally perceived the AI tool as less burdensome, the effect size at this sample size was insufficient to claim a definitive systematic reduction in overall workload. The comparison results are detailed in Table 5.8.

Table 5.8. Paired Comparison of Global Workload Scores (Wilcoxon Signed-Rank Test)

Condition	Mean	SD	Test Statistic	p-value
Manual (Standard)	2.08	1.08	W=295.0	.096
AI-Assisted	1.74	1.03	(z=1.676)	

5.3.2.2 Subscale Analysis: Cognitive vs. Procedural Load

While the global score provides a high-level summary, the granular impact of the AI intervention is best understood by analysing the six individual NASA-TLX subscales. As illustrated in the Radar Chart (Figure 5.2), the AI-assisted workflow demonstrated reductions across all six dimensions, though the magnitude of these reductions varied.

The most substantial benefits were observed in the cognitive domains:

- **Mental Demand** decreased from a mean of **2.13** (Manual) to **1.45** (AI), representing a raw reduction of 0.68. This indicates that the AI tool successfully functioned as a "cognitive off loader," handling the complex pattern recognition tasks that typically tax the clinician.
- **Effort** showed a similar decline, dropping from **2.30** to **1.83** (Δ-0.47), suggesting that clinicians felt they had to work less hard to achieve the same diagnostic result.

In contrast, the reduction in **Performance** anxiety (Δ-0.16) and **Temporal Demand** (Δ-0.23) was less pronounced. This implies that while the tool makes the task *mentally* easier, the procedural aspects, such as the time required to upload images and review predictions—remain comparable to the manual workflow.

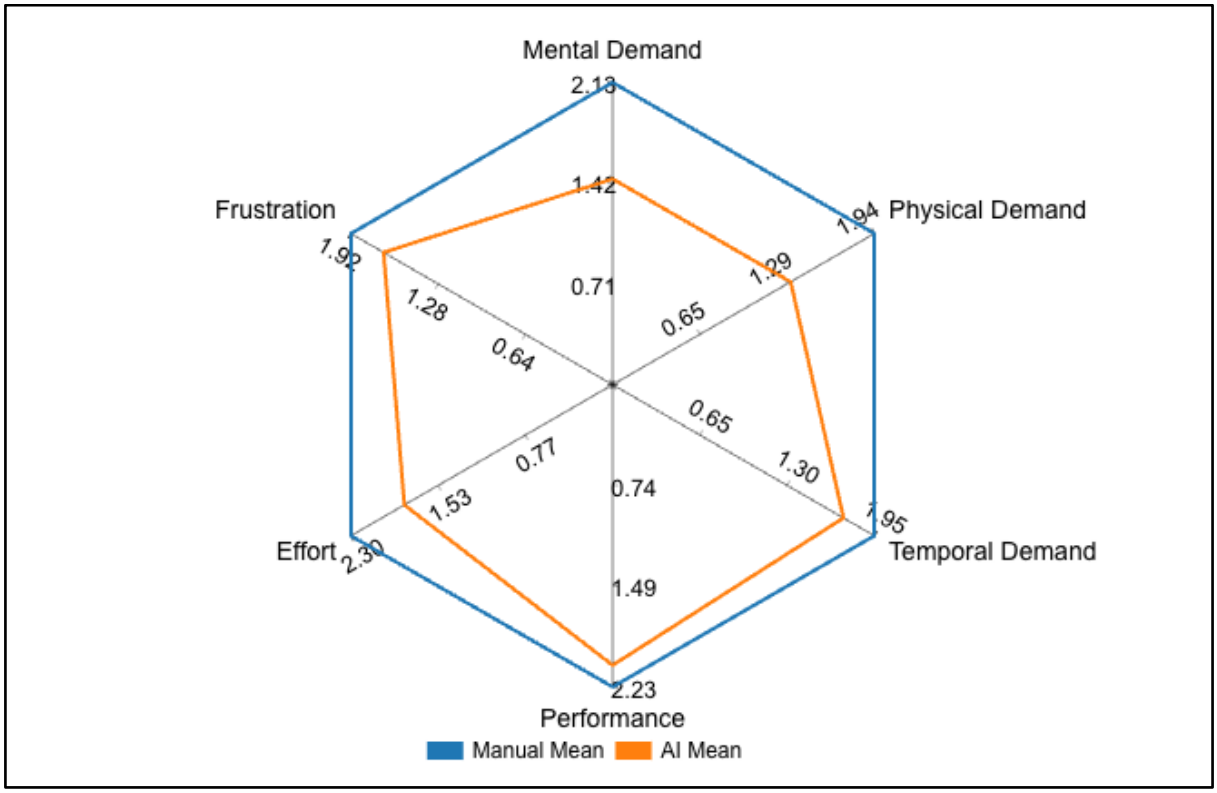


Figure 5.2. Radar chart illustrating the mean scores for the six NASA-TLX subscales. A contraction of the shape (orange area) compared to the baseline (blue area) indicates a reduction in perceived workload, most notably in Mental Demand.

All constructs of the questionnaire showed a positive shift with AI with specific components of the NASA-TLX highlighted key areas of impact. Mental Demand was reported as significantly lower during the AI-assisted task. Temporal demand was also lower, reflecting the potential for AI to reduce time pressure during image interpretation.

Inferential Comparison of Workload Across Practitioner Groups

To determine if the observed workload benefits of the AI-assisted tool varied significantly among different cadres of practitioners, inferential statistics were employed to compare the mean NASA-TLX scores between demographic and professional groups. The difference in raw NASA-TLX scores (Standard Tool workload minus AI-Assisted Tool workload) served as the primary dependent variable to measure the differential impact of the AI tool across groups. It will be referred to as the NASA_DIFF

A Mann-Whitney U test was used to compare the workload reduction scores between male and female participants.

The result showed no statistically significant difference ($p > 0.05$). This finding suggests that there is no significant difference in perceived workload reduction between the genders.

A Kruskal–Wallis test was conducted to examine whether profession (Specialists vs Generalists) influenced the change in workload (NASA_DIFF) between the standard and AI-assisted condition.

Table 5.9 Kruskal Wallis test results for practice years and profession

Factor	Statistic	df	p
Profession	0.001	1	.971
practice_years	1.875	1	.171

Note. Tests are conducted separately for each variable, without accounting for multivariate effects.

The test was not statistically significant, ($H(1) = 0.001$, $p = .971$), suggesting no statistically significant difference in NASA_DIFF between the two groups.

To determine whether participants' years of practice influenced the improvement in workload when using AI assistance, a Kruskal–Wallis H test was performed with NASA_DIFF as the dependent variable and practice-year category as the grouping variable. The results showed that the difference in workload scores between the practice-year groups was not statistically significant, $H(1) = 1.875$, $p = .171$. This suggests that experience level did not play a meaningful role in the extent to which AI assistance reduced perceived workload.

These findings indicate that AI-assisted workload reduction was consistent across experience levels, genders and specialties suggesting that both junior and senior clinicians, as well as experts and non experts derived comparable benefit from the AI system. This may be because of the simpler nature of the interpretation task.

5.3.3 User Acceptance and Usability (TAM 2)

To evaluate the potential for long-term clinical adoption, participant responses were analyzed using the Technology Acceptance Model 2 (TAM 2) framework. Analysis followed the regiment described Chapter 3.8.4 : **Perceived Usefulness (PU)** and **Perceived Ease of Use (PEOU)** were the core constructs analysed. The internal consistency of these constructs was

assessed using Cronbach's Alpha (α), yielding high reliability coefficients for both PU (α= 0.95) and PEOU (α = 0.83), confirming the validity of the instrument for this study.

5.3.3.1 Construct Analysis

Participants expressed strong positive sentiment toward the AI-assisted interface. The results, summarized in Table 5.10, indicate that the system met the critical requirements for deployment in a busy clinical environment.

- **PEOU:** This construct received the highest rating (M = 5.97, SD = 0.9) on a 7-point scale). This strong consensus suggests that the iterative design process—specifically the automation of inference triggers and the simplification of the upload workflow—successfully minimized the learning curve. For generalist doctors operating under time pressure, the tool was perceived as intuitive and low-effort.
- **PU:** The mean score of **5.84** (SD = 1.26) indicates that practitioners viewed the AI outputs not just as novel, but as clinically relevant aids. The slightly higher Standard Deviation compared to PEOU suggests some variability in how different clinicians perceive the value of the advice, likely dependent on their own confidence levels, but the overall trend remains firmly positive.

Table 5.10 Summary of Technology Acceptance Model (TAM 2) Construct Scores

Construct	**Cronbach's α**	**Mean (Max 7)**	**Standard Deviation**	**Interpretation**
Perceived Ease of Use	0.83	5.97	0.9	High Usability
Perceived Usefulness	0.95	5.84	1.26	High Utility

5.3.3.2 Distribution of Sentiment

To visualize the consensus among participants, a diverging stacked bar chart was generated (Figure 5.3). This visualization highlights that negative sentiment (Disagree/Strongly Disagree) was negligible across all constructs. The majority of responses clustered in the 'Agree' and 'Strongly Agree' bands, reinforcing the viability of the tool as a support mechanism in Zambian radiological workflows.

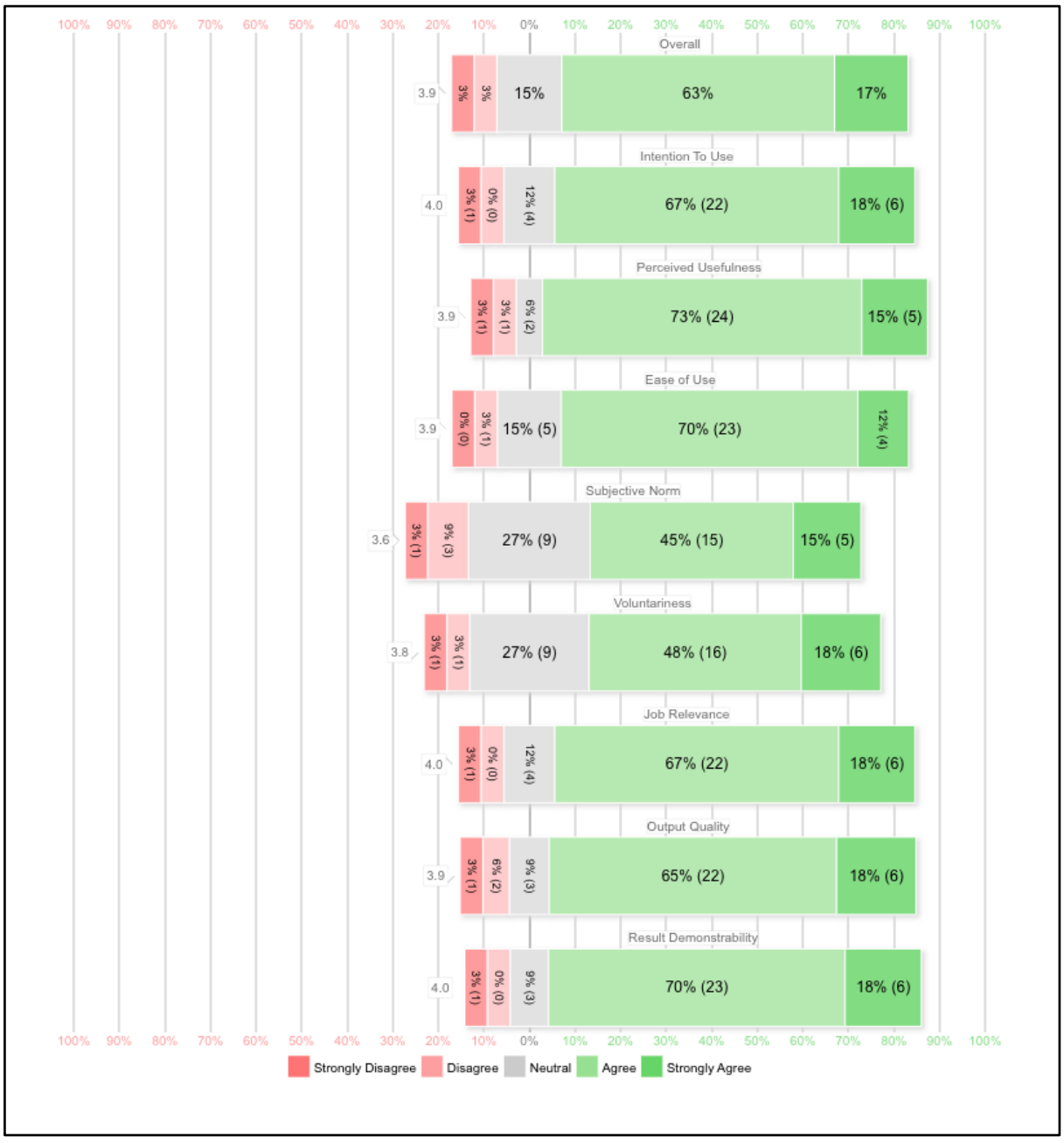


Figure 5.3 Diverging stacked bar chart illustrating the distribution of participant agreement levels across the core TAM 2 constructs. The dominance of the green bars (Agree/Strongly Agree) indicates widespread acceptance

Perceived Usefulness is critical, directly addressing the core research objective of finding AI solutions that are effective for medical practitioners.

Overall, participants reported high perceived usefulness, with mean scores consistently in the agreement range. High PU suggests that the tool addresses documented issues such as high cognitive load and fragmentation of care that hinder timely interpretation. The aggregated mean score for PU was 5.7 (SD=1.3), supporting the feasibility of adopting this type of tool in the local context.

Perceived Ease of Use (PEOU) Perceived Ease of Use relates to the extent a user believes that using the system would be free of effort, which is vital for successful integration into already stressful and resource-limited workflows.

The results indicated strong agreement that the AI tool was easy to use. This positive finding suggests that the interface refinements made following the focus group discussions such as automating the processing workflow on image upload successfully translated into a seamless user experience. PEOU is a strong predictor of technology adoption, particularly when high workload and time pressure are prevalent.

Intention to Use a is a direct measure of the likelihood that practitioners will incorporate the AI solution into their routine work, providing critical input for the selection criteria (Objective 3). Additionally, the Subjective Norms component assessed external social factors influencing the decision to adopt the tool, such as peer recommendation and patient benefit.

Inferential Comparison of Workload Across Practitioner Groups

To determine if the observed workload benefits of the AI-assisted tool varied significantly among different cadres of practitioners, inferential statistics were employed to compare some Tam construct scores between demographic and professional groups.

A Mann-Whitney U test was used to compare the workload reduction scores between male and female participants for PU, PEOU and IU.

Table 5.11 Mann-Whitney U test of Gender comparison stats

Category	U	P
perceived_usefullness	89.00	.664
Intention_to_use	63.00	.104
ease_of_use	124.00	.267

Note. Mann-Whitney U test.

The test found no statistically significant difference between the genders $p > .05$ on any tested construct. This suggests similar experience of the tool between the genders in those areas.

A Kruskal–Wallis test was conducted with the perceived usefulness construct as the dependent variable to examine whether *profession and years of practice* influenced the change in perceived usefulness of the tool.

Table 5.12 Kruskal-Wallis H test of perceived usefulness for practice years and Profession

Factor	Statistic	df	p
practice_years	0.015	1	.901
profession	0.143	1	.705

Note. Tests are conducted separately for each variable, without accounting for multivariate effects.

The results showed no statistically significant difference between profession and practice years.

5.3.4 Qualitative Feedback and Feature Requests

Following the structured quantitative evaluation of workload and usability, participants were invited to provide open-ended qualitative comments regarding their experience with the AI interpretation tool and suggestions for future improvement. These remarks generally fell into three distinct themes: the perceived utility of the tool, necessary constraints regarding professional expertise, and specific suggestions for technical and feature enhancements.

Table 5.13 Experiment Open Response Participant Profile Table

Role	Years of Practice	Participant Code
Medical Doctor	< 1	Medical Doctor 1
Medical Doctor	1 - 5	Medical Doctor 2
Medical Doctor	1 - 5	Medical Doctor 3
Medical Doctor	1 - 5	Medical Doctor 4
Medical Doctor	1 - 5	Medical Doctor 5
Radiologist	5 - 10	Radiologist 1
Radiology Resident	1 - 5	Radiology Resident 1
Radiology Resident	1 - 5	Radiology Resident 2
Radiology Resident	1 - 5	Radiology Resident 3
Radiology Resident	1 - 5	Radiology Resident 4
Radiology Resident	1 - 5	Radiology Resident 5
Radiology Resident	5 - 10	Radiology Resident 6
Radiology Resident	5 - 10	Radiology Resident 7
Pediatric Neurosurgeon	5 - 10	Pediatric Surgeon 1

Perceived Utility and Positive Adoption Sentiment

The majority of comments expressed a highly positive view of the AI tool's potential, focusing on its ability to support and enhance existing diagnostic processes. One key recurring theme was the anticipated increase in clinical confidence and its value as an educational resource. For instance, one participant remarked that ‘The use of AI would increase confidence in diagnosis as well as being a useful teaching tool’ [Medical Doctor 1]. Another simply stated, ‘It's an excellent idea’ [Medical Doctor 2]. Furthermore, the tool was recognized for its potential to improve clinical operations, with a Radiology resident noting, ‘It is a good tool for positive change in patient management’ [Radiology Resident 1].

Professional Constraints and Implementation Context

A significant theme emerged around the need to ensure the AI tool is used appropriately and only by personnel with sufficient medical knowledge to interpret its findings. This sentiment highlighted concerns over deskilling or misuse by non-specialists. A Radiology Resident cautioned that the technology 'Needs to be in the right hands, users should be able to interpret(medical back)'[sic] [Radiology Resident 2]. Similarly, a Radiologist expressed a clear boundary for usage, stating, 'I would only recommend radiology residents and radiologists to use such tools. As other doctors may have challenges'[sic] [Radiologist 1]. Conversely, some comments addressed the necessary flexibility required for specialized users, arguing that 'Interpretation should be in line with radiologist's individual preference' [Radiology Resident 3]. A related constraint noted by a Paediatric surgeon limited the tool's scope, noting they, 'wouldn't recommend for complex imaging. Like CTs' [Paediatric Surgeon 1].

Feature Enhancement and Output Specificity

Participants offered several technical suggestions focusing on improving the depth, localization, and breadth of the AI's output. These responses directly assist in answering Research Question 2. A critical request was to enhance the specificity and explanatory power of the outputs generated by the tool. One Radiology resident suggested 'Improvement in the output text, descriptions of the exact locations of the pathology. Look into more lung and chest conditions' [Radiology Resident 4]. This desire for detailed localization was echoed by a Medical Dr, who proposed that 'The foci for infection/pulmonary infiltration could be better targeted by the AI model' [Medical Doctor 3]. Practitioners also requested that the tool support the full cognitive process of diagnosis, including providing alternative options, suggesting 'For the AI tool to give a suggestion of other diagnoses' [Radiology Resident 5], or that it should '...outline pathologies first and then come up with a conclusion and differentials [Radiology Resident 6].

Regarding input constraints, a General Dr requested the capacity to integrate contextual data: 'There should be a way of inputting the medical details with the x-ray' [Medical Doctor 4]. Concerning the range of pathology covered, there was a 'Strong recommended working with specialised pneumonias'[SIC] [Medical Doctor 5]. Finally, one Radiology resident reflected a more philosophical stance on the rapid changes in the field, stating, 'I don't know where the world is going with technology' [Radiology Resident 7].

5.3 Chapter Summary

This chapter detailed the comprehensive, mixed-methods research methodology employed to investigate, develop, and evaluate an Artificial Intelligence intervention for the Zambian radiological workflow. Grounded in a pragmatic research philosophy, the study adapted the "Research Onion" to frame its mixed-methods approach and utilized CHAT to contextualize AI integration as a complex socio-technical intervention rather than a purely technological one.

The operational execution of the study was systematically guided by the six-phase CRISP-DM lifecycle. Phase I (Business Understanding) utilized a convergent mixed-methods approach, combining a nationwide survey of medical practitioners with semi-structured qualitative interviews, to identify critical workflow bottlenecks and define pneumonia as the high-impact target pathology. Phases II and III outlined the careful selection of proxy datasets and the application of rigorous data augmentation techniques to overcome local data scarcity. In Phase IV (Modeling), a dual-architecture strategy was detailed, employing VGG16 for robust binary classification and Mask R-CNN for explainable object detection.

Phases V and VI bridged the gap between computational modelling and clinical reality. Following an iterative, UCD process driven by expert focus groups, the AI models were integrated into a functional clinical interface. The evaluation of this tool was conducted via a highly controlled, within-subjects quasi-experiment. By utilizing counterbalanced task orderings and validated psychometric instruments (the NASA-TLX for cognitive workload and TAM 2 for technology acceptance), the study ensured that the AI's clinical utility could be objectively measured while controlling for individual diagnostic variance. Finally, the chapter outlined the strict ethical protocols and multi-institutional clearances that governed the handling of medical data and human participants throughout the study.

Chapter 6. Discussion

This chapter interprets the empirical findings presented in Chapter 5, synthesizing the quantitative survey data, technical model metrics, and experimental evaluation results to address the study's core research questions. Rather than merely summarizing the observed outcomes, this discussion analyses why specific workflow breakdowns exist in the Zambian context and *how* the introduction of AI serves as a systemic intervention.

To provide a theoretically grounded interpretation, the findings are examined through the lens of **CHAT** [23, 84] . Within this framework, the radiological workflow is viewed not as a linear technical process, but as a complex **Activity System** where human **Subjects** (clinicians) use **Tools/Artifacts** (imaging hardware, AI) to transform an **Object** (patient images) into an outcome (diagnosis). This chapter identifies critical **systemic** contradictions—such as the tension between high patient volume and limited specialist availability—and evaluates the extent to which the developed AI solution acts as a mediating artifact capable of resolving these tensions.

The discussion is structured as follows:

- **Section 6.1** analyzes the current crisis in radiological workflows, framing inefficiencies as breakdowns between the Community and the Rules of the activity system.
- **Section 6.2** evaluates the AI tool as a mediating artifact, specifically examining its role in resolving workload contradictions for generalist practitioners.
- **Section 6.3** provides a vigorous technical defense of the model's performance trade-offs (High Recall vs. Low Precision) within the specific safety context of resource-constrained healthcare.
- **Section 6.4** discusses the shifting Division of Labor and the ethical imperatives for professional oversight and user competence.

6.1 The Crisis of 'Object' and 'Community': Systemic Breakdowns in Radiological Workflows

The first objective of this study was to empirically characterize the challenges inherent in Zambian radiological image interpretation. The findings from the contextual inquiry (Section 5.1) confirm that the 'workforce shortage' is not merely a staffing metric, but a fundamental constraint that distorts the entire activity system of diagnosis. The survey data revealed a severe imbalance where generalist medical doctors, who comprise most of the interpretation workforce, operate with low confidence in complex modalities and under significant time pressure.

When viewed through the lens of Cultural-Historical Activity Theory (CHAT), these challenges represent deep-seated **systemic contradictions**. The primary contradiction exists between the **Subject** (the overburdened generalist clinician) and the **Object** (the high volume and complexity of radiological images). This tension is exacerbated by the scarcity of specialized **Tools** (diagnostic hardware) and a breakdown in the **Community** interactions (information flow between referring and interpreting doctors).

Consistent with global findings on radiological capacity in the Global South and Zambia, where scarcity compromises timely and accurate diagnosis [32, 40, 41, 44, 119]. the operational reality in Zambia is characterized by a 'survival workflow'—a system designed to manage volume rather than optimize diagnostic precision. The following subsections analyse two specific dimensions of this crisis: the rupture in clinical information flow and the limitation of physical infrastructure.

6.1.1 The 'Analog' Gap: Information Flow as a Community Breakdown

A critical finding from the clinician survey was that **74.5 percent** of respondents agreed their diagnostic effectiveness is significantly reduced when patient clinical history is incomplete or missing. While missing clinical data is a known challenge in radiology globally, in the Zambian context, this statistic represents more than a logistical oversight; it signifies a fundamental breakdown in the **Community** node of the activity system.

In a functional radiological workflow, the **Subject** (radiologist/interpreter) and the **Community** (referring clinicians, nurses) operate in a collaborative loop where the image and the patient's history are synthesized to form a diagnosis. However, the qualitative findings (Theme 2) reveal that this loop is severed by the reliance on 'analogue' transfer mechanisms for digital data. The current **Rules** of the system dictate that information travels physically—via handwritten request forms carried by patients or CDs manually transported between departments. This creates a single point of failure: if the paper form is lost or illegible, the interpreter is left to operate in a 'diagnostic void,' isolated from the clinical context required to differentiate pathologies with similar radiographic presentations.

This fragmentation contradicts the core objective of radiological practice. As noted by the [RaRe] participant in the interviews, many pathologies are 'mimickers' on static imaging; without the narrative context provided by the Community (e.g., 'cough for 3 weeks' vs. 'acute trauma'), the **Object** of the activity remains ambiguous. This forces the clinician to rely solely on visual pattern recognition, a strategy that is inherently prone to error and lower confidence, as evidenced by the 'Confidence-Competence Mismatch' reported in Section 5.1.1.2.

This finding aligns with broader literature on health informatics in LMICs, where the lack of integrated RIS or EHR creates "data silos" [65]. Unlike high-resource settings where the **Division of Labor** is supported by automated data merging, Zambian practitioners must expend significant cognitive effort attempting to reconstruct the patient's story from fragmented sources. Consequently, the barrier to effective diagnosis is not merely the *scarcity* of interpreters, but the *opacity* of the information they are asked to interpret.

6.1.2 Infrastructure as a Limiting Artifact: The Impact of Non-Dedicated Hardware

Beyond information gaps, the physical **Tools** available for interpretation constitute a second layer of systemic failure. In Activity Theory, tools are 'mediating artifacts' that shape how the Subject interacts with the Object. Ideally, high-resolution, calibrated monitors serve as transparent tools that reveal the true nature of the pathology. However, the qualitative data (Theme 1) exposes a critical 'Tool–Object' contradiction: the widespread reliance on non-optimized hardware actively obscures the diagnostic object.

Interview participants vividly described a workflow where ‘secondary capture’—viewing photographs of X-ray films on personal mobile phones or computer screens with poor luminance—is a normalized workaround. This adaptation, while resourceful, fundamentally degrades the **Object** (the medical image). The artifacts introduced by this process, such as glare, geometric distortion, and loss of contrast resolution, create ‘pseudo-pathologies’ that can mislead the interpreter. As noted by one [SRMO], this forces the clinician to question whether a visual opacity is a true lesion or merely a reflection on the screen.

This infrastructure deficit significantly impacts diagnostic confidence. When the mediating tool is unreliable, the Subject must expend additional mental effort to filter out technical noise, increasing cognitive load and the likelihood of error. This finding resonates with Karera et al. [47], who identified limited equipment and suboptimal viewing conditions as primary barriers to effective service delivery in similar African radiological settings.

The implication for this study is profound: any AI intervention introduced into this environment must be robust enough to handle ‘noisy’ inputs. Furthermore, it suggests that AI tools should ideally include image enhancement or quality assurance features (as requested in Section 5.3.4) to compensate for the degradation caused by the physical infrastructure. By validating the image quality before interpretation, the AI can serve as a stabilizing artifact, mitigating the risks imposed by the limitations of the existing hardware.

6.2 The AI Tool as a "Mediating Artifact": Resolving Workload Contradictions

The second objective of this study was to develop and evaluate an AI solution capable of supporting medical practitioners in this constrained environment. In the language of CHAT, the introduction of the AI-assisted interface represents the deployment of a new **Mediating Artifact** intended to resolve the critical contradiction between the **Subject** (the generalist clinician operating under high pressure) and the **Object** (the high volume of complex chest radiographs).

The experimental evaluation results (Section 5.3) provide empirical evidence that this intervention was successful in altering the nature of the diagnostic activity. By offloading specific cognitive tasks to the algorithm, the tool reshaped the ‘Cost’ of the activity for the

clinician. The following subsections analyse the specific mechanisms of this workload reduction and discuss how the tool serves as a scaffold for generalist competence.

6.2.1 Cognitive Offloading: Resolving the Subject-Object Tension

The most significant finding from the workload evaluation was the substantial reduction in **Mental Demand** (Raw Score Δ -0.68) and Effort (Δ -0.47) reported by participants when using the AI-assisted workflow. In the context of activity theory, this reduction signals that the AI tool successfully functioned as an instrument of cognitive offloading [82].

Radiological interpretation is a dual-process task involving *visual search* (finding the lesion) and decision making (classifying the pathology). For the generalist clinician, the 'search' phase is cognitively expensive, particularly when fatigue or low confidence is a factor. The bounding box visualizations generated by the Mask R-CNN model acted as an attentional cue, effectively automating the visual search component. By directing the clinician's gaze immediately to 'suspicious regions,' the tool reduced the cognitive resources required to scan the image, allowing the Subject to reallocate mental energy toward the higher-level task of clinical correlation.

This interpretation helps explain the high Perceived Usefulness scores (Mean PU: 5.7). The tool did not merely provide a label; it reduced the "decision anxiety" associated with the fear of missing a subtle finding. By providing a probabilistic "second opinion," the AI resolved the tension between the Subject's limited capacity and the Object's complexity. This finding mirrors results from similar studies in emergency medicine, where AI triage tools have been shown to lower the cognitive burden on non-specialist staff by filtering out normal cases and highlighting abnormalities [80, 96]

6.2.2 Scaffolding for Generalists: Bridging the Competence Gap

The survey results (Section 5.1.1.2) identified a critical 'Confidence-Competence Mismatch,' where 73 percent of generalists interpreting complex imaging reported low diagnostic confidence. This finding reveals a vulnerability in the healthcare system: the Division of Labor demands that generalists perform specialist tasks without the requisite specialist support. The AI tool addresses this gap by functioning as a 'Zone of Proximal Development' (ZPD) scaffold [108].

In Vygotskian theory (the root of CHAT), the ZPD represents what a learner can achieve with guidance that they cannot yet achieve alone. By providing immediate feedback (e.g., "78% Probability of Pneumonia" and specific localization), the AI tool extends the diagnostic capability of the generalist doctor. It serves as a synchronous 'digital mentor,' validating their hunches or flagging overlooked pathologies in real-time.

This scaffolding effect explains why the workload reduction was consistent across both Specialists and Generalists (Section 5.3.2). For generalists, the benefit is primarily educational and reassuring, the tool acts as the 'senior colleague' they lack in rural settings. This aligns with the 'Optimistic Pragmatism' theme found in the qualitative feedback, where participants explicitly valued the tool as a 'learning aid' that could increase their confidence. By creating a feedback loop where none existed before, the AI tool does not just automate the task; it upskills the workforce, potentially mitigating the long-term impact of the radiologist shortage [100].

6.3 The Safety Philosophy of 'Recall over Precision': Technical Trade-offs in Context

While the AI tool demonstrated high user acceptance and workload reduction, the technical performance metrics (specifically the low Precision (15%) relative to high Recall (79%))require rigorous clinical justification. In a pure computer vision context, a low precision score might indicate a suboptimal model. However, in the specific context of a resource-constrained medical screening workflow, this performance profile reflects a deliberate safety philosophy.

The design of the AI intervention prioritized **Recall (Sensitivity)** over Precision (Positive Predictive Value). The following subsections defend this trade-off by analyzing the relative clinical cost of False Positives versus False Negatives in the Zambian healthcare environment.

6.3.1 The Clinical Rationale for High Recall (Sensitivity)

In the Zambian public health sector, the scarcity of radiologists means that a chest radiograph interpreted by a generalist is often the *only* imaging assessment a patient receives before treatment decisions are made. In this high-stakes environment, the cost of a **False Negative**

(missing a true pathology) is catastrophic. A missed pneumonia diagnosis can lead to sepsis, untreated tuberculosis spread, or death, outcomes that are irreversible.

Consequently, the AI model was optimized to maximize **Recall (79%)**. A high recall ensures that the system functions as a highly sensitive 'net,' capturing the vast majority of abnormal cases. As highlighted by the qualitative findings (Theme 2), 'Time is brain'. Delays in specialist reporting (up to two weeks for some modalities) mean that clinicians cannot wait for a 'gold standard' report to catch missed cases. Therefore, the AI must ensure that no sick patient is inadvertently discharged or ignored. By prioritizing sensitivity, the tool aligns with the ethical imperative of **non-maleficence** (do no harm) by minimizing the risk of under-diagnosis in a setting where 'second looks' are operationally impossible.

This approach is consistent with international guidelines for screening tools, particularly in Global Health contexts (e.g., TB triage), where the World Health Organization (WHO) explicitly prioritizes high sensitivity ($\geq$ 90% target) over specificity to ensure effective case finding [113]. While our model's 79 percent recall falls slightly short of the ideal WHO target for automated reading, it significantly outperforms the 'zero support' baseline currently available to rural doctors.

6.3.2 Defending Low Precision: AI as a 'Triage Tool' vs. 'Gold Standard'

The trade-off for high recall was a **precision of 15 percent**, meaning the model generated a significant number of **False Positives** (flagging benign areas as suspicious). While technically inefficient, clinical safety theory dictates that in a screening workflow, a False Positive is a 'safe fail-state.'

When the AI incorrectly flags a region (e.g., a rib crossing mistaken for an opacity), the consequence is that the clinician scrutinizes that area. This induced vigilance forces the human **Subject** to engage with the image, verify the finding, and ultimately dismiss it using their clinical judgment. The 'cost' is merely a few seconds of additional attention. In contrast, a False Negative would provide false reassurance, leading to the premature closure of the diagnostic process.

Therefore, the low precision is acceptable because the AI is not designed as an **autonomous diagnostician** (a 'Gold Standard' replacement), but as a **triage and alert system**. Its role is to say, 'Look here,' not 'This is definitely pneumonia.' This distinction is critical. As established in the literature on 'AI Safety Nets' [76], systems designed for human-in-the-loop workflows should be tuned to provoke human verification rather than passive acceptance. By accepting over-flagging, the system ensures that the human doctor remains the final arbiter, preserving professional agency while mitigating the risk of missed diagnoses. This operational logic transforms the 'technical flaw' of low precision into a 'clinical feature' of heightened vigilance.

6.4 Division of Labor and Professional Guardrails: The Ethics of Implementation

The introduction of AI into the Zambian radiological ecosystem does not merely change *how* diagnosis happens; it fundamentally shifts who is empowered to perform it. In CHAT terms, this represents a renegotiation of the **Division of Labor**. Traditionally, definitive image interpretation is the exclusive domain of the Radiologist. However, the operational reality (0.5 radiologists per million people) has already forced a de facto shift, placing the burden on generalist medical officers.

The proposed AI solution validates and formalizes this shifted role, equipping generalists with specialist-level pattern recognition capabilities. However, this empowerment brings significant ethical risks. The qualitative feedback (Section 5.3.4) highlighted a strong professional concern that AI could lead to 'deskilling' or dangerous over-reliance by untrained staff. The following subsections discuss the implications of this role shift and define the necessary 'Professional Guardrails' to ensure safe implementation.

6.4.1 Shifting Roles: The Generalist as the New Primary Interpreter

The deployment of the AI tool effectively flattens the traditional hierarchy of radiological practice. By providing generalist doctors with immediate, quantified 'second opinions,' the system legitimizes their role as primary interpreters. In the pre-AI workflow, a generalist interpreting a chest X-ray does so as a provisional measure, often fraught with uncertainty and the 'hidden workload' of ad-hoc peer consultation (Section 5.1.2.3). With the AI tool, this activity is scaffolded, transforming the generalist from a passive *requester* of reports into an active *producer* of preliminary diagnostic data.

This shift in the **Division of Labor** is critical for the sustainability of the Zambian health system. It moves the model from a 'bottleneck' architecture (where all images must wait for a radiologist) to a 'distributed' architecture (where only complex or ambiguous cases require specialist escalation). This aligns with the concept of 'task shifting' promoted by the WHO for resource-constrained settings[34]. However, unlike traditional task shifting which relies on human training alone, this model uses the AI artifact to maintain quality standards across a distributed workforce.

6.4.2 The 'User Competence' Criterion: Preventing Deskilling

While empowering generalists is necessary, the study's qualitative findings (Section 5.3.4) revealed a sharp professional anxiety regarding competence. Participants explicitly warned that the tool 'needs to be in the right hands' and expressed concern that junior staff might 'turn off their brains' and blindly accept the AI's probability score. This highlights a new contradiction: the risk of **automation bias**, where the human Subject cedes agency to the Tool.

To mitigate this, a critical criterion for implementation must be established: **User Competence and Training**. The AI tool cannot be deployed as a standalone 'black box.' Instead, its introduction must be accompanied by new **Rules** governing its use. Specifically, users must be trained not just on *how* to use the interface, but on the *limitations* of the model (e.g., its low precision). They must understand that the AI is a screening aid, not a diagnostic authority.

This aligns with Morley et al.'s [68] ethical framework for AI in healthcare, which emphasizes that the ultimate moral and legal responsibility for a diagnosis must remain with the human clinician. Therefore, the 'Division of Labor' must be codified: the AI proposes, but the competent human disposes. Establishing this "competence criterion" prevents the technology from becoming a crutch that leads to deskilling, ensuring instead that it remains a tool for capacity building.

[34] https://www.unaids.org/sites/default/files/media_asset/ttr_taskshifting_en_0.pdf

6.5 Limitations

While this study successfully demonstrates the potential of AI to mediate radiological workflows in Zambia, several limitations must be acknowledged to contextualize the findings.

6.5.1 Architectural Trade-offs and Model Justification

The selection of VGG16 for this study, rather than newer architectures like ViTs or EfficientNet, is driven primarily by dataset size constraints and training computation considerations. While recent studies demonstrate that ViTs offer high theoretical efficiency and exhibit greater robustness to image perturbations compared to traditional CNNs [9], they fundamentally lack the spatial inductive biases inherent to convolutional architectures. Consequently, ViTs require massive-scale datasets (such as JFT-300M or ImageNet-21k) to converge and generalize effectively [19].

In specialized domains like chest radiography, where available annotated datasets are comparatively limited in size, CNNs retain a distinct advantage. Comparative evaluations in medical imaging have shown that traditional CNN architectures consistently outperform or match Transformers on small- to medium-sized datasets, largely due to their ability to maximize performance through effective transfer learning and weight initialization [62].

Furthermore, it is acknowledged that VGG16 is computationally heavy due to its large parameter count (approximately 138 million parameters). As discussed in Section 6.1.2 regarding non-dedicated hardware, this large footprint poses notable limitations for deployment on edge devices or mobile platforms. However, the actual computational resources required to fine-tune the convolutional layers of a pre-trained VGG16 model on a small medical dataset are generally lower than the heavy computational overhead required to train the self-attention mechanisms of Transformers. For a server-side inference model where data efficiency, training stability, and comparative benchmarking against existing literature are paramount, the trade-off strongly favors VGG16. Its selection is predicated on establishing a well-documented foundation that aligns with the constraints of the available data and computational resources.

6.5.2 Training Data and Algorithmic Bias

A significant limitation involves the data used to train the model. The AI models were trained primarily on public datasets sourced from non-African populations. Although data augmentation techniques were employed to simulate local image quality degradation, the lack of a large-scale, locally annotated Zambian dataset raises legitimate concerns regarding algorithmic bias. It remains possible that the model's performance may vary when encountering pathologies with region-specific radiographic presentations, such as advanced pulmonary tuberculosis or HIV-associated opportunistic infections.

6.5.3 Clinical Scope and Outcome Measurement

Furthermore, the technical scope was restricted to a **s**ingle pathology (Pneumonia) and a single modality (Chest X-Ray). While justified by the high burden of respiratory disease in Zambia, this narrow focus limits the immediate generalizability of the findings to the broader radiological workload, which includes complex trauma (fractures) and cross-sectional imaging (CT/MRI) where the shortage of specialists is even more acute. Additionally, while the study measured perceived workload and utility, it did not measure direct clinical outcomes. We established that doctors felt less burdened and *believed* the tool was useful, but we did not track longitudinal metrics such as 'reduction in patient waiting time' or 'change in mortality rates.' These downstream impacts are the ultimate test of clinical efficacy and require a longer-term implementation study to validate.

6.5.4 Experimental Design and Ecological Validity

Finally, the study faces limitations regarding its experimental setup and participant demographics. The evaluation experiment was controlled, whereas real-world clinics in Zambia may experience power outages, network issues, and other infrastructure factors not matched in the experiment; therefore, the results represent a best-case scenario. Additionally, the participant population was largely radiology experts (34%) and urban doctors. This means these results might not apply directly to a Clinical Officer in a rural facility. The sample size for the evaluation (N=47), though sufficient for a pilot usability study, is relatively small for broad inferential generalization. The statistical trends observed (such as the workload reduction ($p = .096$)) suggest real effects that might have reached statistical significance with a larger

cohort. Future studies should aim to expand the participant pool to include a wider range of rural facilities to fully capture the heterogeneity of the Zambian health infrastructure.

Chapter 7. Conclusion

This thesis opened with a stark characterization of the radiological landscape in Zambia: a system defined by a critical shortage of specialists, where generalist clinicians operate in an "informal" diagnostic capacity, overburdened by volume and isolated from expert support. The primary objective of this research was not merely to engineer a technical artifact, but to investigate whether AI could serve as a systemic intervention to resolve these operational contradictions.

By applying CHAT as an analytical lens, this study has demonstrated that the crisis in Zambian radiology is not just a resource gap but a breakdown in the activity system itself, specifically, a rupture in the community information flow and a tension between the subject's limited capacity and the object's complexity. The results presented in the preceding chapters confirm that a locally adaptable, human-in-the-loop AI model can effectively mediate these tensions. When designed with high sensitivity and integrated seamlessly into clinical workflows, AI transforms the generalist doctor from an unsupported interpreter into a scaffolded diagnostician, thereby stabilizing the delivery of care in resource-constrained settings.

7.1 Synthesis of Findings: Resolving the Crisis of Capacity

The empirical evidence gathered in this study paints a detailed picture of a healthcare system operating at the limits of its capacity. The initial contextual inquiry revealed that the challenges in Zambian radiology extend beyond the mere headcount of radiologists. The profound ‘confidence-competence mismatch’ (where 73 percent of generalists interpret complex imaging despite low confidence) signals a systemic risk to patient safety. This risk is exacerbated by ‘analogue’ information gaps, with nearly three-quarters of practitioners lacking the clinical history required for effective diagnosis, and infrastructural deficits that force doctors to view high-stakes medical images on non-dedicated mobile devices.

Against this backdrop, the developed AI solution proved to be a potent mediating artifact. By prioritizing **Recall (79%)** over Precision, the technical design explicitly aligned with the clinical imperative of the ‘safety net,’ ensuring that potential pathologies are flagged for review rather than missed. This design choice, though technically modest compared to ‘gold standard"\’ benchmarks, proved highly effective in the local context. The experimental

evaluation demonstrated that the tool successfully reduced the cognitive burden of interpretation, lowering the **Mental Demand** on practitioners and acting as an educational scaffold.

Critically, the high acceptance of the tool (Perceived Usefulness: 5.7) suggests that the 'cultural resistance' often cited in literature is not an immutable barrier. When an AI tool is positioned not as an autonomous replacement but as a 'spark plug' for human reasoning—and when it is integrated directly into the workflow (e.g., via Orthanc plugins)—it resolves the paralysis of the overburdened clinician. The synthesis of these findings confirms the central thesis: in a resource-limited environment, the value of AI lies less in its ability to outperform a radiologist, and more in its ability to empower a generalist to function safely in the radiologist's absence.

7.2 Major Contributions and Significance

This research contributes to the body of knowledge on AI in Global Health by moving beyond model development to systemic validation. The major contributions are:

1. **Contextualized Evidence Base:** The study provides analyses of the 'secondary capture' phenomenon (using phones for diagnosis) and the 'confidence gap' among Zambian generalists. This data fills a critical gap in the global understanding of radiological bottlenecks in Southern Africa.
2. **Validated 'Safety Net' Prototype:** The research demonstrates the feasibility of deploying open-source, locally adaptable AI architectures (Mask R-CNN on commodity hardware) that prioritize sensitivity over specificity. This technical artifact proves that high-performance computing is not a prerequisite for high-impact clinical support.
3. **A Framework for AI-Enabled Task Shifting:** Perhaps most significantly, this work offers a blueprint for national health policy. By validating that generalist doctors can effectively use AI to scaffold their decision-making, the study suggests that **AI-Enabled Task Shifting** is a viable national strategy for the Ministry of Health. Rather than waiting decades to train sufficient radiologists, the health system can leverage such tools to formally empower and certify generalists to handle routine interpretation, reserving the scarce specialist workforce for complex, high-priority

cases. This represents a paradigm shift from a ‘referral-dependent’ model to a ‘distributed-competence’ model of care.

4. **Evaluation Framework**: It introduces a combined NASA-TLX and TAM2 evaluation approach for assessing AI’s impact on clinical workload and usability.
5. **Application of CHAT**: It applies Cultural-Historical Activity Theory to uncover systemic contradictions and guide responsible integration of AI in clinical practice.

7.3 Recommendations for Policy and Practice

Based on the systemic contradictions identified in this study, the following recommendations are proposed for health administrators, the Ministry of Health, and policy stakeholders:

1. **Infrastructure First:** Before large-scale AI deployment, there is an urgent need to invest in basic digital infrastructure. The finding that practitioners rely on mobile phones for diagnosis highlights a critical gap. Hospitals should prioritize the procurement of adequate viewing monitors and local PACS networks. AI cannot correct for the loss of diagnostic information caused by ‘secondary capture’ on low-resolution screens.
2. **Mandatory ‘AI Competence’ Training:** To mitigate the risks of ‘automation bias’ and deskilling, the introduction of AI tools must be accompanied by a formalized training curriculum. This training should not focus merely on software operation but on **AI literacy**—teaching clinicians to understand probability scores, recognize false positives, and maintain their own diagnostic agency. Certification in ‘AI-Assisted Interpretation’ should be a prerequisite for access to these tools.
3. **Redefining Liability Frameworks:** The regulatory environment must evolve to clarify the legal status of AI predictions. Policy should explicitly define AI as a **triage agent** rather than a diagnostic authority. Liability for the final report must remain with the human clinician, but the system should legally recognize the ‘validity’ of a generalist seeking AI consultation, thereby protecting doctors who use these tools in good faith to bridge the specialist gap.

7.4 Recommendations for Future Work

This study has opened a door to the integration of AI in Zambian radiology, but many questions remain unanswered. To deepen this inquiry, the research community should consider the following directions:

1. **Longitudinal Clinical Impact:** While this study established *perceived* utility, the ultimate test is patient outcome. Future research should ask: *Does the introduction of AI triage actually reduce the time-to-treatment for pneumonia patients in rural clinics?* A longitudinal study tracking patient outcomes over 6–12 months is the necessary next step.
2. **Expanding the Pathology Scope:** Pneumonia is a critical starting point, but the burden of disease includes tuberculosis, cardiac enlargement, and trauma. Future technical work should explore: Can multi-label models be trained on local datasets to handle the "polypathology" typical of Zambian patients (e.g., TB-HIV co-infection)?
3. **The Ethics of Dependence:** As AI becomes more capable, a sociological question arises: How does long-term reliance on AI scaffolding affect the development of "clinical intuition" in junior doctors? Research is needed to monitor the learning curves of medical interns to ensure that the tool functions as a teacher, not a crutch.

7.5 Concluding Remarks

At the outset of this thesis, the image of Zambian radiology was one of scarcity: a solitary radiologist facing an impossible workload, while patients waited weeks for answers that might come too late. The journey of this research has shown that while we cannot clone the radiologist, we can clone their expertise. By embedding specialist-level pattern recognition into the tools used by everyday doctors, we do not just improve a workflow; we democratize access to diagnostic certainty. The AI tool developed here is not a replacement for the human expert, but a promise that in the future, no clinician in Zambia will have to face a complex diagnosis alone.

Chapter 8. References

[1] Addressing the Limitations of Medical Data in AI: 2024. *https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/addressing-limitations-medical-data-ai*. Accessed: 2025-05-15.

[2] Aggarwal, A. et al. 2023. Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. *Journal of medical internet research*. 25, (Feb. 2023), e40789. DOI:https://doi.org/10.2196/40789.

[3] Alaran, M.A. et al. 2025. Challenges and opportunities of artificial intelligence in African health space. *Digital health*. 11, (Jan. 2025), 20552076241305915. DOI:https://doi.org/10.1177/20552076241305915.

[4] AlQudah, A.A. et al. 2021. Technology acceptance in healthcare: A systematic review. *Applied Sciences (Basel, Switzerland)*. 11, 22 (Nov. 2021), 10537. DOI:https://doi.org/10.3390/app112210537.

[5] Arambula, A.M. and Bur, A.M. 2020. Ethical considerations in the advent of artificial intelligence in otolaryngology. *Otolaryngology--head and neck surgery: official journal of American Academy of Otolaryngology-Head and Neck Surgery*. 162, 1 (Jan. 2020), 38–39. DOI:https://doi.org/10.1177/0194599819889686.

[6] ARTIFICIAL INTELLIGENCE (AI) IN ZAMBIA: 2023. *https://bongohive.co.zm/artificial-intelligence-ai-in-zambia/*. Accessed: 2025-05-15.

[7] Artificial Intelligence and Machine Learning in Software as a Medical Device: 2025. *https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#:~:text=improve%20its%20performance.-,How%20Is%20the%20FDA%20Considering%20Regulation%20of%20Artificial%20Intelligence%20and,Novo%20classification%2C%20or%20premarket%20approval.* Accessed: 2025-05-15.

[8] Baek, G. et al. 2025. AI chatbots for psychological health for health professionals: Scoping review. *JMIR human factors*. 12, 1 (Mar. 2025), e67682. DOI:https://doi.org/10.2196/67682.

[9] Bhojanapalli, S. et al. 2021. Understanding robustness of Transformers for image classification. *arXiv [cs.CV]*.

[10] Bishop, C.M. 2006. *Pattern recognition and machine learning*. Springer.

[11] Braun, V. and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*. 3, 2 (Jan. 2006), 77–101. DOI:https://doi.org/10.1191/1478088706qp063oa.

[12] Bwanga, O. et al. 2020. Need for image reporting by radiographers in Zambia. *Medical journal of Zambia*. 46, 3 (Jan. 2020), 215–220. DOI:https://doi.org/10.55320/mjz.46.3.223.

[13] Bwanga, O. and Chanda, E. 2020. Challenges in Radiation Protection in Healthcare: A Case of Zambia. 2, 1 (Feb. 2020), 7–14. DOI:https://doi.org/10.36349/EASJRIT.2020.v02i01.002.

[14] Chiweza, C. et al. 2024. Can artificial intelligence-augmented fetal monitoring prevent intrapartum stillbirth and neonatal death in a low-income setting: An observational study? *BJOG: an international journal of obstetrics and gynaecology*. 131, 1 (Jan. 2024), 109–111. DOI:https://doi.org/10.1111/1471-0528.17321.

[15] CIDRZ Partners with Qure.ai for AI TB Screening: *https://www.qure.ai/news_press_coverages/centre-for-infectious-disease-research-in-zambia-cidrz-partners-with-qure-ai-for-tb-screening*. Accessed: 2025-05-15.

[16] Ciecierski-Holmes, T. et al. 2022. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *npj digital medicine*. 5, 1 (Oct. 2022), 162. DOI:https://doi.org/10.1038/s41746-022-00700-y.

[17] Clark, M. and Bailey, S. 2024. *Chatbots in health care: Connecting patients to information: Emerging health technologies*. Canadian Agency for Drugs and Technologies in Health.

[18] Deng, J. et al. 2009. ImageNet: A large-scale hierarchical image database. *2009 IEEE*

Conference on Computer Vision and Pattern Recognition (Jun. 2009), 248–255.
[19] Dosovitskiy, A. et al. 2020. *An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR)*.
[20] Dunn, O.J. 1964. Multiple comparisons using rank sums. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*. 6, 3 (Aug. 1964), 241. DOI:https://doi.org/10.2307/1266041.
[21] Dwork, C. et al. 2012. Fairness through awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (New York, NY, USA, Jan. 2012).
[22] Edemekong, P.F. et al. 2025. Health Insurance Portability and Accountability Act (HIPAA) compliance. *StatPearls*. StatPearls Publishing.
[23] Engestrom, Y. 2014. *Learning by expanding: An activity-theoretical approach to developmental research*. Cambridge University Press.
[24] Enterprise Medical Imaging for Streamlined Radiological Diagnosis in Zambian Public Health Facilities: 2021. *The University of Zambia. https://doi.org/10.13140/RG.2.2.11215.61603*. Accessed: 2023-08-01.
[25] EUR-Lex - 02017R0745-20250110 - EN - EUR-Lex: *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02017R0745-20250110*. Accessed: 2025-05-15.
[26] Evaluation Methods for Artificial Intelligence (AI)-Enabled Medical Devices: Performance Assessment and Uncertainty Quantification: 2024. *https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/evaluation-methods-artificial-intelligence-ai-enabled-medical-devices-performance-assessment-and*. Accessed: 2025-05-15.
[27] Fu, Y. et al. 2021. A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica: PM: An International Journal Devoted to the Applications of Physics to Medicine and Biology: Official Journal of the Italian Association of Biomedical Physics (AIFB)*. 85, (May 2021), 107–122. DOI:https://doi.org/10.1016/j.ejmp.2021.05.003.
[28] Garrucho, L. et al. 2022. Domain generalization in deep learning based mass detection in mammography: A large-scale multi-center study. *Artificial Intelligence in Medicine*. 132, 102386 (Oct. 2022), 102386. DOI:https://doi.org/10.1016/j.artmed.2022.102386.
[29] Grier, R.A. 2015. How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Human Factors and Ergonomics Society. Annual Meeting*. 59, 1 (Sep. 2015), 1727–1731. DOI:https://doi.org/10.1177/1541931215591373.
[30] Gu, H. et al. 2021. Lessons Learned from Designing an AI-Enabled Diagnosis Tool for Pathologists. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (Apr. 2021), 1–25. DOI:https://doi.org/10.1145/3449084.
[31] Hart, S.G. and Staveland, L.E. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*. P.A. Hancock and N. Meshkati, eds. Elsevier. 139–183.
[32] Hawkins, J. 2001. Addressing the shortage of radiologists. *Radiology management*. 23, 4 (Jul. 2001), 26–28.
[33] He, K. et al. 2015. Deep residual learning for image recognition. *arXiv [cs.CV]*.
[34] He, K. et al. 2017. Mask R-CNN. *arXiv [cs.CV]*.
[35] Hesamian, M.H. et al. 2019. Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of Digital Imaging*. 32, 4 (Aug. 2019), 582–596. DOI:https://doi.org/10.1007/s10278-019-00227-x.
[36] Hicks, S.A. et al. 2022. On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*. 12, 1 (Apr. 2022). DOI:https://doi.org/10.1038/s41598-022-09954-8.
[37] Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 6, 2 (1979), 65–70.
[38] Hoofnagle, C.J. et al. 2019. The European Union general data protection regulation: what it is and what it means. *Information & communications technology law*. 28, 1 (Jan. 2019), 65–98. DOI:https://doi.org/10.1080/13600834.2019.1573501.

[39] Hosny, A. et al. 2018. Artificial intelligence in radiology. *Nature Reviews Cancer*. 18, 8 (2018), 500–510.
[40] Hricak, H. et al. 2021. Medical imaging and nuclear medicine: a Lancet Oncology Commission. *The lancet oncology*. 22, 4 (Apr. 2021), e136–e172. DOI:https://doi.org/10.1016/S1470-2045(20)30751-8.
[41] Jing, A.B. et al. 2025. AI solutions to the radiology workforce shortage. *npj health systems*. 2, 1 (May 2025), 20. DOI:https://doi.org/10.1038/s44401-025-00023-6.
[42] Jodogne, S. 2018. The orthanc ecosystem for medical imaging. *Journal of Digital Imaging*. 31, 3 (Jun. 2018), 341–352. DOI:https://doi.org/10.1007/s10278-018-0082-y.
[43] Johnson, A. et al. 2024. MIMIC-CXR Database. PhysioNet.
[44] Kalidindi, S. and Gandhi, S. 2023. Workforce Crisis in Radiology in the UK and the Strategies to Deal With It: Is Artificial Intelligence the Saviour? *Cureus*. 15, 8 (Aug. 2023), e43866. DOI:https://doi.org/10.7759/cureus.43866.
[45] Kallianos, K. et al. 2019. How far have we come? Artificial intelligence for chest radiograph interpretation. *Clinical Radiology*. 74, 5 (May 2019), 338–345. DOI:https://doi.org/10.1016/j.crad.2018.12.015.
[46] Kamal, S.A. et al. 2020. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). *Technology in Society*. 60, 101212 (Feb. 2020), 101212. DOI:https://doi.org/10.1016/j.techsoc.2019.101212.
[47] Karera, A. et al. 2024. Operational challenges and collaborative solutions in radiology image interpretation: perspectives from imaging departments in a low-resource setting. *Journal of medical radiation sciences*. 71, 4 (Dec. 2024), 564–572. DOI:https://doi.org/10.1002/jmrs.815.
[48] Karera, A. et al. 2024. Radiology image interpretation services in a low-resource setting: Medical doctors' experiences and the potential role of radiographers. *Radiography (London, England: 1995)*. 30, 2 (Mar. 2024), 560–566. DOI:https://doi.org/10.1016/j.radi.2024.01.009.
[49] Kaushik, A. et al. 2025. Challenges and opportunities for data sharing related to artificial intelligence tools in health care in low- and middle-income countries: Systematic review and case study from Thailand. *Journal of medical internet research*. 27, 1 (Feb. 2025), e58338. DOI:https://doi.org/10.2196/58338.
[50] Kelly, C.J. et al. 2019. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*. 17, 1 (Dec. 2019). DOI:https://doi.org/10.1186/s12916-019-1426-2.
[51] Kermany, D. et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning Cell. 172, 1122–1131.
[52] Kermany, D. 2018. Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification. Mendeley.
[53] Kingma, D.P. and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv [cs.LG]*.
[54] Lång, K. et al. 2023. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *The Lancet Oncology*. 24, 8 (Aug. 2023), 936–944. DOI:https://doi.org/10.1016/s1470-2045(23)00298-x.
[55] Litjens, G. et al. 2017. A survey on deep learning in medical image analysis. *arXiv [cs.CV]*.
[56] Liu, F. et al. 2019. Fully Automated Diagnosis of Anterior Cruciate Ligament Tears on Knee MR Images by Using Deep Learning. *Radiology. Artificial intelligence*. 1, 3 (May 2019), 180091. DOI:https://doi.org/10.1148/ryai.2019180091.
[57] Liu, X. et al. 2019. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*. 1, 6 (Oct. 2019), e271–e297. DOI:https://doi.org/10.1016/s2589-7500(19)30123-2.
[58] Liu, Y. et al. 2020. Rethinking computer-aided tuberculosis diagnosis. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Jun. 2020).
[59] Maeda, Y. et al. 2019. Fully automated diagnostic system with artificial intelligence using

endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointestinal endoscopy*. 89, 2 (Feb. 2019), 408–415. DOI:https://doi.org/10.1016/j.gie.2018.09.024.
[60] Majkowska, A. et al. 2020. Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. *Radiology*. 294, 2 (Feb. 2020), 421–431. DOI:https://doi.org/10.1148/radiol.2019191293.
[61] Makasa, E. 2009. The Human Resource crisis in the Zambian Health Sector – a discussion paper. *Medical journal of Zambia*. 35, 3 (Oct. 2009). DOI:https://doi.org/10.4314/mjz.v35i3.46522.
[62] Matsoukas, C. et al. 2021. Is it time to replace CNNs with transformers for medical images? *arXiv [cs.CV]*.
[63] Mbewe, C. et al. 2020. An audit of licenced Zambian diagnostic imaging equipment and personnel. *The Pan African medical journal*. 36, (May 2020), 32. DOI:https://doi.org/10.11604/pamj.2020.36.32.21043.
[64] Methods and Tools for Effective Postmarket Monitoring of Artificial Intelligence (AI)-Enabled Medical Devices: 2024. *https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/methods-and-tools-effective-postmarket-monitoring-artificial-intelligence-ai-enabled-medical-devices*. Accessed: 2025-05-15.
[65] Mollura, D.J. et al. 2020. Artificial intelligence in low- and middle-income countries: Innovating global health radiology. *Radiology*. 297, 3 (Dec. 2020), 513–520. DOI:https://doi.org/10.1148/radiol.2020201434.
[66] Momani, A. 2025. *Artificial Intelligent Implications on Health Data Privacy and Confidentiality*.
[67] Morid, M.A. et al. 2021. A scoping review of transfer learning research on medical image analysis using ImageNet. *Computers in Biology and Medicine*. 128, 104115 (Jan. 2021), 104115. DOI:https://doi.org/10.1016/j.compbiomed.2020.104115.
[68] Morley, J. et al. 2020. The ethics of AI in health care: A mapping review. *Social science & medicine (1982)*. 260, 113172 (Sep. 2020), 113172. DOI:https://doi.org/10.1016/j.socscimed.2020.113172.
[69] Morley, J. and Floridi, L. 2020. An ethically mindful approach to AI for health care. *Lancet*. 395, 10220 (Jan. 2020), 254–255. DOI:https://doi.org/10.1016/S0140-6736(19)32975-7.
[70] Morley, J. and Floridi, L. 2024. The ethics of AI in health care: An updated mapping review.
[71] Murdoch, B. 2021. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC medical ethics*. 22, 1 (Sep. 2021), 122. DOI:https://doi.org/10.1186/s12910-021-00687-3.
[72] Ng'andwe, E.M. and Bwanga, O. 2022. Factors affecting the ability of radiographers to deliver imaging services in rural parts of Zambia. *Radiography (London, England: 1995)*. 28, 3 (Aug. 2022), 758–765. DOI:https://doi.org/10.1016/j.radi.2022.03.005.
[73] Nguyen, H.Q. et al. 2022. VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations. *Scientific Data*. 9, 1 (Jul. 2022), 429. DOI:https://doi.org/10.1038/s41597-022-01498-w.
[74] Nigatu, A.M. et al. 2023. Medical imaging consultation practices and challenges at public hospitals in the Amhara regional state, Northwest Ethiopia: a descriptive phenomenological study. *BMC health services research*. 23, 1 (Jul. 2023), 787. DOI:https://doi.org/10.1186/s12913-023-09652-9.
[75] Nordling, L. 2019. A fairer way forward for AI in health care. *Nature*. 573, 7775 (Sep. 2019), S103–S105. DOI:https://doi.org/10.1038/d41586-019-02872-2.
[76] Oakden-Rayner, L. et al. 2020. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. *Proceedings of the ACM Conference on Health, Inference, and Learning*. 2020, (Apr. 2020), 151–159. DOI:https://doi.org/10.1145/3368555.3384468.
[77] Performance Evaluation Methods for Evolving Artificial Intelligence (AI)-Enabled Medical Devices: 2024. *https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/performance-evaluation-methods-evolving-artificial-*

intelligence-ai-enabled-medical-devices. Accessed: 2025-05-15.
[78] Popa, S.L. et al. 2021. Non-Alcoholic Fatty Liver Disease: Implementing Complete Automated Diagnosis and Staging. A Systematic Review. *Diagnostics (Basel, Switzerland)*. 11, 6 (Jun. 2021). DOI:https://doi.org/10.3390/diagnostics11061078.
[79] the Precise4Q consortium et al. 2020. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*. 20, 1 (Dec. 2020). DOI:https://doi.org/10.1186/s12911-020-01332-6.
[80] Rajpurkar, P. et al. 2017. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv [cs.CV]*.
[81] Regulatory Evaluation of New Artificial Intelligence (AI) Uses for Improving and Automating Medical Practices: 2024. *https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/regulatory-evaluation-new-artificial-intelligence-ai-uses-improving-and-automating-medical-practices*. Accessed: 2025-05-15.
[82] Risko, E.F. and Gilbert, S.J. 2016. Cognitive offloading. *Trends in Cognitive Sciences*. 20, 9 (Aug. 2016), 676–688. DOI:https://doi.org/10.1016/j.tics.2016.07.002.
[83] Roth, W.-M. 2007. Vygotsky's neglected legacy": Cultural-historical activity theory. *Review of Educational Research*. 77, (2007), 186–232.
[84] Roth, W.-M. 2007. Vygotsky's neglected legacy": Cultural-historical activity theory. *Review of Educational Research*. 77, (2007), 186–232.
[85] RSNA Pneumonia Detection Challenge: *https://www.kaggle.com/competitions/rsna-pneumonia-detection-challenge/data*. Accessed: 2025-12-18.
[86] Rubio, S. et al. 2004. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. *Psychologie appliquee [Applied Psychology]*. 53, 1 (Jan. 2004), 61–86. DOI:https://doi.org/10.1111/j.1464-0597.2004.00161.x.
[87] Said, S. et al. 2020. Validation of the raw National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire to assess perceived workload in patient monitoring tasks: Pooled analysis study using mixed models. *Journal of Medical Internet Research*. 22, 9 (Sep. 2020), e19472. DOI:https://doi.org/10.2196/19472.
[88] Saito, T. and Rehmsmeier, M. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One*. 10, 3 (Mar. 2015), e0118432. DOI:https://doi.org/10.1371/journal.pone.0118432.
[89] Saunders, M. et al. 2023. *Research methods for business students*. Pearson Education.
[90] Shapiro, S.S. and Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*. 52, 3-4 (Dec. 1965), 591–611. DOI:https://doi.org/10.1093/biomet/52.3-4.591.
[91] Shawa, A. and Phiri, L. 2025. Deployment and evaluation of intelligent DICOM viewers in low-resource settings: Orthanc plugin for semi-automated interpretation of medical images. *Communications in Computer and Information Science*. Springer Nature Switzerland. 134–143.
[92] Shiraishi, J. et al. 2000. Development of a digital image database for chest radiographs with and without a lung nodule. *AJR. American Journal of Roentgenology*. 174, 1 (Jan. 2000), 71–74. DOI:https://doi.org/10.2214/ajr.174.1.1740071.
[93] Simonyan, K. and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv [cs.CV]*.
[94] Society for Imaging 2009. *Practical Imaging Informatics: Foundations and Applications for PACS Professionals*. Springer Science & Business Media.
[95] Sokolova, M. and Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*. 45, 4 (Jul. 2009), 427–437. DOI:https://doi.org/10.1016/j.ipm.2009.03.002.
[96] Stewart, J. et al. 2024. Attitudes towards artificial intelligence in emergency medicine. *Emergency medicine Australasia: EMA*. 36, 2 (Apr. 2024), 252–265. DOI:https://doi.org/10.1111/1742-6723.14345.
[97] Tahir, M.Y. et al. 2022. A review of teleradiology in Africa - Towards mobile teleradiology in

Nigeria. *SA journal of radiology*. 26, 1 (Jan. 2022), 2257. DOI:https://doi.org/10.4102/sajr.v26i1.2257.
[98] Tajbakhsh, N. et al. 2017. Convolutional neural networks for medical image analysis: Full training or fine tuning? *arXiv [cs.CV]*.
[99] Takona, J.P. 2023. Research design: qualitative, quantitative, and mixed methods approaches / sixth edition. *Quality & Quantity*. (Nov. 2023). DOI:https://doi.org/10.1007/s11135-023-01798-2.
[100] Topol, E.J. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 25, 1 (Jan. 2019), 44–56. DOI:https://doi.org/10.1038/s41591-018-0300-7.
[101] U-Net Architecture Explained: 2023. *https://www.geeksforgeeks.org/u-net-architecture-explained/*. Accessed: 2024-07-15.
[102] U.S. Food and Drug Administration: 2025. *https://www.fda.gov/*. Accessed: 2025-05-15.
[103] Vayena, E. et al. 2018. Machine learning in medicine: Addressing ethical challenges. *PLoS medicine*. 15, 11 (Nov. 2018), e1002689. DOI:https://doi.org/10.1371/journal.pmed.1002689.
[104] Venkatesh, V. et al. 2003. User acceptance of information technology: Toward A unified View1. *MIS Quarterly*. 27, 3 (Sep. 2003), 425–478. DOI:https://doi.org/10.2307/30036540.
[105] Venkatesh, V. and Davis, F.D. 2000. A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*. 46, 2 (Feb. 2000), 186–204. DOI:https://doi.org/10.1287/mnsc.46.2.186.11926.
[106] Victor, A. 2025. Artificial intelligence in global health: An unfair future for health in Sub-Saharan Africa? *Health affairs scholar*. 3, 2 (Feb. 2025). DOI:https://doi.org/10.1093/haschl/qxaf023.
[107] Vygotsky, L.S. 1980. *Mind in society*. Harvard University Press.
[108] Vygotsky, L.S. 1978. *Mind in society: The development of higher psychological processes*. Harvard University Press.
[109] Wang, X. et al. ChestX-ray14 Dataset.
[110] Wang, X. et al. 2017. *ChestX-ray14: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases*.
[111] Weiner, E.B. et al. 2025. Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice. *PLOS digital health*. 4, 4 (Apr. 2025), e0000810. DOI:https://doi.org/10.1371/journal.pdig.0000810.
[112] What are Convolutional Neural Networks? 2021. *https://ibm.com/topics/convolutional-neural-networks*. Accessed: 2024-07-15.
[113] WHO consolidated guidelines on tuberculosis: module 2: screening: systematic screening for tuberculosis disease: 2021. *https://www.who.int/publications/i/item/9789240022676*. Accessed: 2025-12-31.
[114] Yakubu, M. et al. 2025. Themed challenges to solve data scarcity in Africa: A proposition for increasing local data collection and integration. *arXiv [cs.CY]*.
[115] Yan, C. et al. 2022. A fully automatic artificial intelligence-based CT image analysis system for accurate detection, diagnosis, and quantitative severity evaluation of pulmonary tuberculosis. *European radiology*. 32, 4 (Apr. 2022), 2188–2199. DOI:https://doi.org/10.1007/s00330-021-08365-z.
[116] Zambia National Health Strategic Plan 2017-2021: *https://www.medbox.org/document/zambia-national-health-strategic-plan-2017-2021*. Accessed: 2025-09-01.
[117] Zech, J.R. et al. 2018. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Medicine*. 15, 11 (Nov. 2018), e1002683. DOI:https://doi.org/10.1371/journal.pmed.1002683.
[118] Zulu, E.O. and Phiri, L. 2022. Enterprise medical imaging in the global south: Challenges and opportunities. *2022 IST-Africa Conference (IST-Africa)* (May 2022).
[119] 2022. Radiology Workforce Shortage in Europe. *Insights into Imaging*. 13, 1 (2022).

DOI:https://doi.org/10.1186/s13244-022-01206-9.
[120] 2022. *Republic of Zambia Ministry of Health: Annual Health Statistical Report 2017 - 2021: Tracking Performance of the NHSP 2017 - 2021. Republic of Zambia Ministry of Health*.

Appendices

Appendix A: Semi Structured Interview Guide

Intro

- Say Hello and Introduce yourself.
- Allow the participant to introduce themselves

Overview

Thank you for taking time to participate in this study. You will be asked a series of questions that will help create solutions for the challenges faced by medical practitioners in medical image interpretation. The questions will be around understanding your workflow, challenges you experience and what requirements you would have for AI solutions to assist you in your work.

Understanding Workflow

- Can you describe your typical workflow when it comes to image interpretation in your practice?
- What tools or technologies do you currently use for image interpretation and how effective do you find them?

Understanding Challenges

- What are the major challenges you face in interpreting medical images?
- How do these challenges impact your efficiency and workload?
- How much time do you spend researching your suspected findings?
- What limitations do you find in the technology you use in interpretation?
- What impact does that have on your accuracy and efficiency?

Potential Requirements

- If we could automate a part of your interpreting process which would it be?
- Assuming we created a specialized interpretation research assistant for you, how would you like it to work?
- Assuming we could assist with the interpretation of x-rays, what pathologies would you want assistance with interpretation for?

Appendix C: NASA Task Load Index

How mentally demanding was the task? *

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

How physically demanding was the task? *

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

How hurried or rushed was the pace of the task? *

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

How successful were you in accomplishing what you were asked to do? *

	1	2	3	4	5	6	7	8	9	10	
Very High	○	○	○	○	○	○	○	○	○	○	Very Low

How hard did you have to work to accomplish your level of performance? *

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

How insecure, discouraged, irritated, stressed, and annoyed were you? *

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

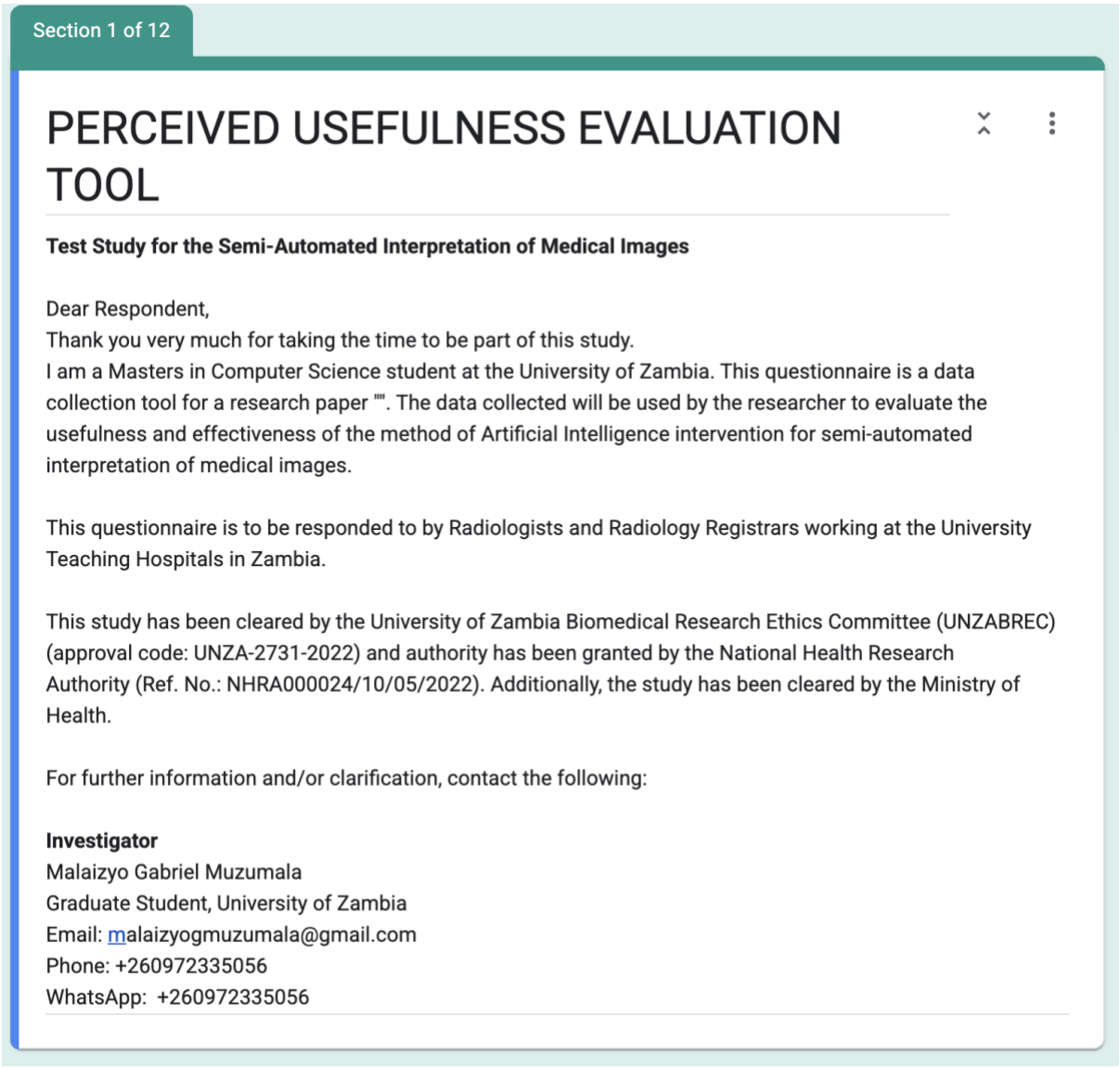
Figure Chapter-1. Chapter-1.

Appendix D: In person Observation Plan

Points of Particular interest

- Turnaround time for reading of specific suspected pathologies in specific modalities.
- Most help requested modality to see which one is most challenging.
- General Observation and keen listening to suggestions.

Appendix E: Perceived Usefulness and Ease of use Questionnaire



Section 1 of 12

PERCEIVED USEFULNESS EVALUATION TOOL

Test Study for the Semi-Automated Interpretation of Medical Images

Dear Respondent,
Thank you very much for taking the time to be part of this study.
I am a Masters in Computer Science student at the University of Zambia. This questionnaire is a data collection tool for a research paper "". The data collected will be used by the researcher to evaluate the usefulness and effectiveness of the method of Artificial Intelligence intervention for semi-automated interpretation of medical images.

This questionnaire is to be responded to by Radiologists and Radiology Registrars working at the University Teaching Hospitals in Zambia.

This study has been cleared by the University of Zambia Biomedical Research Ethics Committee (UNZABREC) (approval code: UNZA-2731-2022) and authority has been granted by the National Health Research Authority (Ref. No.: NHRA000024/10/05/2022). Additionally, the study has been cleared by the Ministry of Health.

For further information and/or clarification, contact the following:

Investigator
Malaizyo Gabriel Muzumala
Graduate Student, University of Zambia
Email: malaizyogmuzumala@gmail.com
Phone: +260972335056
WhatsApp: +260972335056

Section 2 of 12

Research Ethics Committee Clearance Certificate

Description (optional)

Image title



UNIVERSITY OF ZAMBIA
BIOMEDICAL RESEARCH ETHICS COMMITTEE

Telephone: +260 977925304
Telegrams: UNZA, LUSAKA
Telex: UNZALU ZA 44370
Fax: + 260-1-250753
Federal Assurance No. FWA00000338

Ridgeway Campus
P.O. Box 50110
Lusaka, Zambia
E-mail: unzarec@unza.zm
IRB00001131 of IORG0000774

5th May, 2022

Your REF. No. 2731-2022

Section 3 of 12

Grant of Authority to Conduct Research

Description (optional)

Image title



NATIONAL HEALTH RESEARCH AUTHORITY
Paediatric Centre of Excellence, University Teaching Hospital, P.O. Box 30075, LUSAKA
Chalala Office Lot No. 18961/M, Off Kasama Road, P.O. Box 30075, LUSAKA
Tell: +260211 250309 | Email: znhrasec@nhra.org.zm | www.nhra.org.zm

Ref No: NHRA000024/10/05/2022 **Date: 10th May, 2022**

Section 4 of 12

Participant Information

Description (optional)

Enter your participant number

Short answer text

Section 5 of 12

Intention to Use

Description (optional)

Assuming you had access to the Artificial Intelligence to automatically interpret images, would you use it? *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Would you recommend other Doctors to use the Artificial Intelligence to automatically interpret images? *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Section 6 of 12

Percieved Usefulness

Description (optional)

Do you think using the Artificial Intelligence to automatically interprete images, would improve your performance or work rate at your place of work. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Do you feel using the Artificial Intelligence to automatically interprete images, would bring about effectiveness in regards to your line of work? *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Would you find the Artificial Intelligence to automatically interpret images, to be a useful tool in your line of work? *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly disagree

Section 7 of 12

Percieved Ease of Use

Description (optional)

My interaction with the Artificial Intelligence to automatically interpret images, was clear and understandable. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Interacting with the Artificial Intelligence to automatically interpret images, did not require a lot of mental or physical effort. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I would find the Artificial Intelligence to automatically interpret images, to be easy to use. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I would find it easy to get the Artificial Intelligence to automatically interpret images, to do what I want it to do. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Section 8 of 12

Subjective Norm

Description (optional)

My supervisors would allow me to use the Artificial Intelligence to automatically interpret images. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

My patients would benefit from my use of the Artificial Intelligence to automatically interpret images. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Section 9 of 12

Voluntariness

Description (optional)

My use of the Artificial Intelligence to automatically interpret images, would be voluntary. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

My supervisor would not require me to use the Artificial Intelligence to automatically interpret images. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Section 10 of 12

Job Relevance

Description (optional)

In my job, usage of the Artificial Intelligence to automatically interpret images, would be important. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

In my job, usage of Artificial Intelligence to automatically interprete images, would be relevant. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Section 11 of 12

Output Quality

Description (optional)

I had no problem with the output quality of the Artificial Intelligence to automatically interpret images. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

The quality of the output I would get from the Artificial Intelligence to automatically interpret images would be high. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Section 12 of 12

Result Demonstrability

Description (optional)

The results of using the Artificial Intelligence to automatically interpret images would be positive to me. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I believe I could communicate to others the benefits of using the Artificial Intelligence to automatically interpret images. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

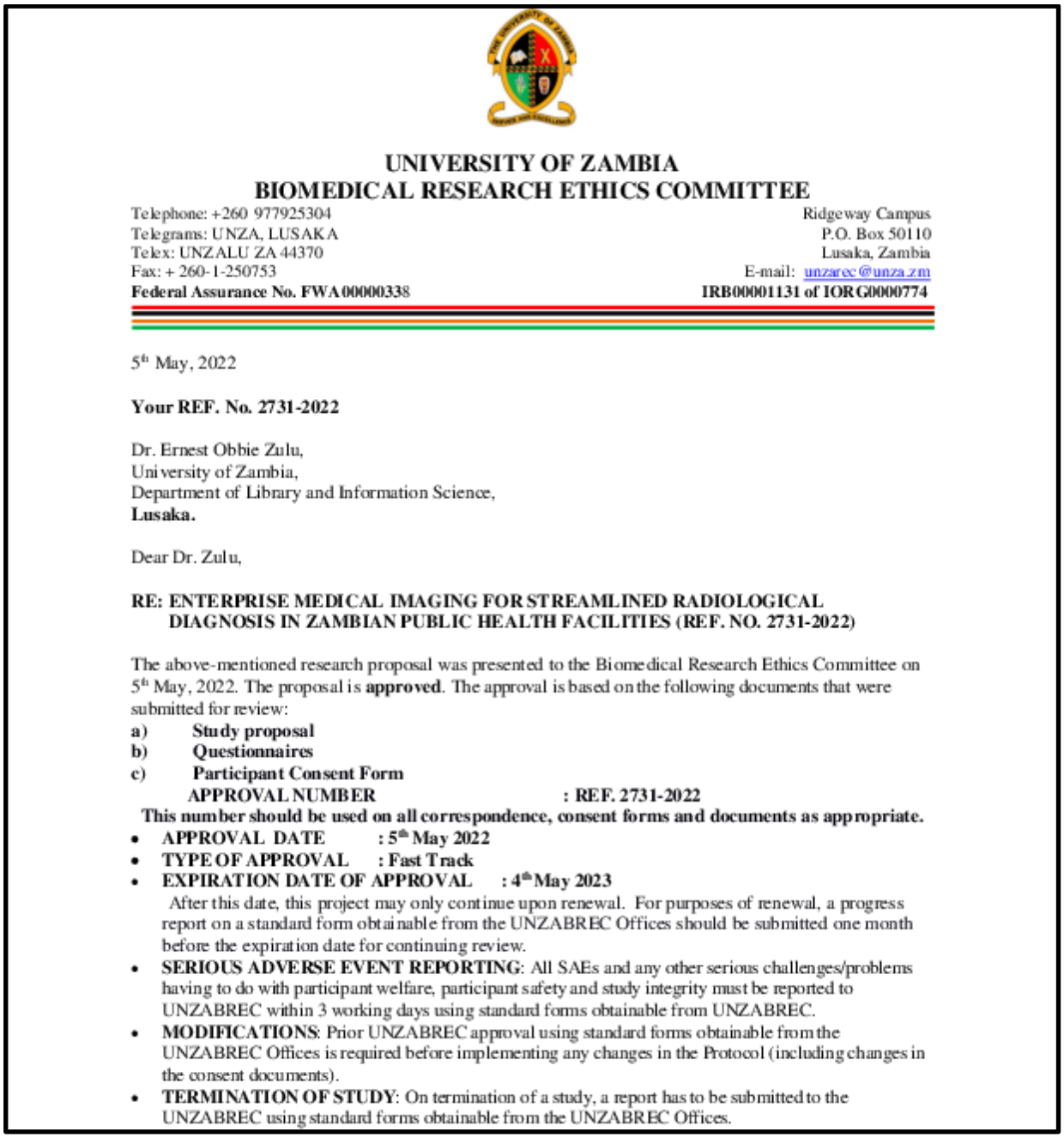
I would have no difficulty telling others about the results of using the Artificial Intelligence to automatically interpret images. *

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

Do you have any comments and/or suggestions on the Artificial Intelligence to automatically interpret images, for future improvements? *

Short answer text

Appendix F: University of Zambia Biomedical Research Ethics Committee Approval



UNIVERSITY OF ZAMBIA
BIOMEDICAL RESEARCH ETHICS COMMITTEE

Telephone: +260 977925304
Telegrams: UNZA, LUSAKA
Telex: UNZALU ZA 44370
Fax: + 260-1-250753
Federal Assurance No. FWA00000338

Ridgeway Campus
P.O. Box 50110
Lusaka, Zambia
E-mail: unzarec@unza.zm
IRB00001131 of IORG0000774

5th May, 2022

Your REF. No. 2731-2022

Dr. Ernest Obbie Zulu,
University of Zambia,
Department of Library and Information Science,
Lusaka.

Dear Dr. Zulu,

RE: ENTERPRISE MEDICAL IMAGING FOR STREAMLINED RADIOLOGICAL DIAGNOSIS IN ZAMBIAN PUBLIC HEALTH FACILITIES (REF. NO. 2731-2022)

The above-mentioned research proposal was presented to the Biomedical Research Ethics Committee on 5th May, 2022. The proposal is **approved**. The approval is based on the following documents that were submitted for review:

a) **Study proposal**
b) **Questionnaires**
c) **Participant Consent Form**

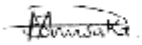
APPROVAL NUMBER : REF. 2731-2022

This number should be used on all correspondence, consent forms and documents as appropriate.

- **APPROVAL DATE : 5th May 2022**
- **TYPE OF APPROVAL : Fast Track**
- **EXPIRATION DATE OF APPROVAL : 4th May 2023**
  After this date, this project may only continue upon renewal. For purposes of renewal, a progress report on a standard form obtainable from the UNZABREC Offices should be submitted one month before the expiration date for continuing review.
- **SERIOUS ADVERSE EVENT REPORTING**: All SAEs and any other serious challenges/problems having to do with participant welfare, participant safety and study integrity must be reported to UNZABREC within 3 working days using standard forms obtainable from UNZABREC.
- **MODIFICATIONS**: Prior UNZABREC approval using standard forms obtainable from the UNZABREC Offices is required before implementing any changes in the Protocol (including changes in the consent documents).
- **TERMINATION OF STUDY**: On termination of a study, a report has to be submitted to the UNZABREC using standard forms obtainable from the UNZABREC Offices.

- **NHRA:** You are advised to obtain final study clearance and approval to conduct research in Zambia from the National Health Research Authority (NHRA) before commencing the research project.
- **QUESTIONS**: Please contact the UNZABREC on Telephone No. +260977925304 or by e-mail on unzarec@unza.zm.
- **OTHER**: Please be reminded to send in copies of your research findings/results for our records. You are also required to submit electronic copies of your publications in peer-reviewed journals that may emanate from this study. Use the online portal: unza.rhinno.net for further submissions.

Yours sincerely,



Sody Mweetwa Munsaka, BSc., MSc., PhD
CHAIRPERSON
Tel: +260977925304
E-mail: s.munsaka@unza.zm

Appendix G: National Health Research Authority Ethical Clearance Approval

NATIONAL HEALTH RESEARCH AUTHORITY
Paediatric Centre of Excellence, University Teaching Hospital, P.O. Box 30075, LUSAKA
Chalala Office Lot No. 18961/M, Off Kasama Road, P.O. Box 30075, LUSAKA
Tell: +260211 250309 | Email: znhrasec@nhra.org.zm | www.nhra.org.zm

Ref No: NHRA000024/10/05/2022 **Date: 10th May, 2022**

The Principal Investigator,
Ernest Obbie Zulu,
University of Zambia
Lusaka, Zambia.

Dear Ernest Obbie Zulu,

Re: Request for Authority to Conduct Research

The National Health Research Authority is in receipt of your request for authority to conduct research titled **"Enterprise Medical Imaging for Streamlined Radiological Diagnosis in Zambian Public Health Facilities."**

I wish to inform you that following submission of your request to the Authority, our review of the same and in view of the ethical clearance, this study has been **approved** on condition that:

1. The relevant Provincial and District Medical Officers where the study is being conducted are fully appraised;
2. Progress updates are provided to NHRA quarterly from the date of commencement of the study;
3. The final study report is cleared by the NHRA before any publication or dissemination within or outside the country;
4. After clearance for publication or dissemination by the NHRA, the final study report is shared with all relevant Provincial and District Directors of Health where the study was being conducted, University leadership, and all key respondents.

Yours sincerely,

Prof. Godfrey Biemba
Director/CEO
National Health Research Authority

Appendix H: Ministry of Health Ethical Clearance Approval



All Correspondence should be addressed to the Permanent Secretary
Telephone: +260 211 253640/5
Fax: +260 211 253344

REPUBLIC OF ZAMBIA
MINISTRY OF HEALTH

In reply please quote:
MOH/
No.....

NDEKE HOUSE
P. O. BOX 30205
LUSAKA

16th May, 2022

Obbie Zulu
LUSAKA

RE: REQUEST FOR AUTHORITY TO CONDUCT RESEARCH

Reference is made to your letter dated 25th April, 2022 in which you requested the Ministry, for permission to conduct a research titled "***Enterprise Medical Imaging for Streamlined Radiological Diagnosis in Zambia Public Health Facilities***". I wish to inform you that my office has no objection to this request provided that;

1. The relevant Institution Director where the study is being conducted are fully appraised;
2. The final study report is cleared by NHRA before any publication or dissemination within or outside the country; and
3. After clearance for publication or dissemination by NHZRA, the final study report is shared with the Ministry.

Kindly ensure minimum interruption in health service delivery to the selected health you will undertake your research.

By copy of this letter, the Provincial, District Health Offices and facilities are advised to allow you undertake the above mentioned research and provide you with the relevant support.

Yours faithfully

Prof. Lackson Kasonka
Permanent Secretary- Technical Services
MINISTRY OF HEALTH

Appendix I: University Teaching Hospitals Ethical Clearance Approval (1/2)



REPUBLIC OF ZAMBIA
MINISTRY OF HEALTH
University Teaching Hospitals -Adult

Fax: +260 211 250305
e-mail: oduth@yahoo.com

P/Bag Rw 1X
Lusaka - Zambia
Tel: +260 211 253942 (Switch Board)
+260 211 251451

OFFICE OF THE SENIOR MEDICAL SUPERINTENDENT

Our Ref:

Your Ref:

5th September, 2022

Dr. Ernest Obbie Zulu
University of Zambia
Department of Library & Information Science
P O Box 50110
LUSAKA

Dear Dr. Zulu,

RE: REQUEST FOR AUTHORITY TO CONDUCT RESEARCH

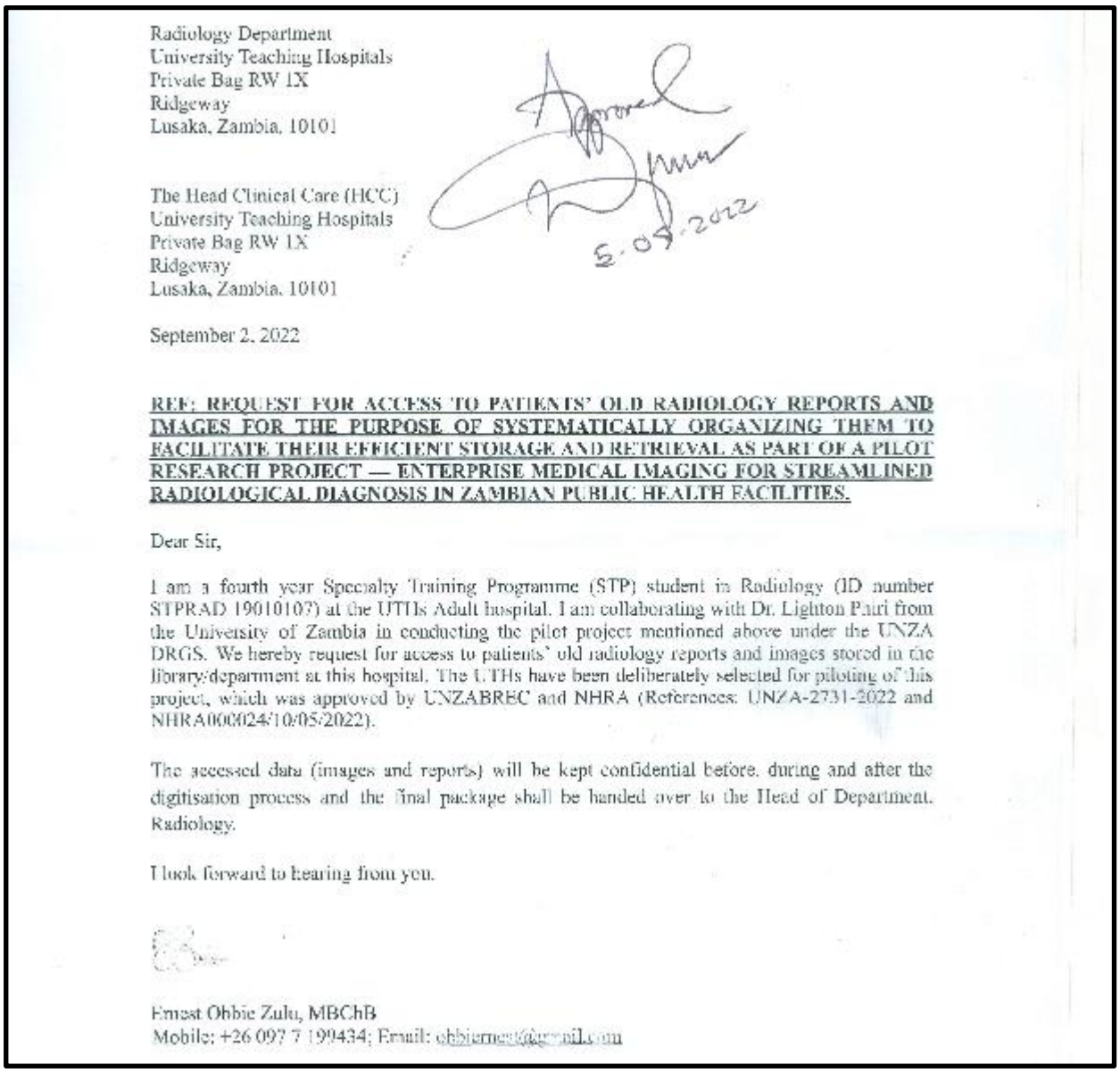
The University Teaching Hospital – Adult is in receipt of your letter dated 5th September, 2022 in which you had requested to conduct a research titled ***"Enterprise Medical Imaging for Streamlined Radiological Diagnosis in Zambia Public Health Facilities" at the University Teaching Hospital."***

I wish to inform you that permission has been granted and you are advised to liaise with the Head of Department.

Yours faithfully,

Dr. Mwila Lunasha
Head Clinical Care
for/Senior Medical Superintendent
UNIVERSITY TEACHING HOSPITALS - ADULT

Appendix J: University Teaching Hospitals Ethical Clearance Approval (2/2)



Radiology Department
University Teaching Hospitals
Private Bag RW 1X
Ridgeway
Lusaka, Zambia, 10101

Approved
5.09.2022

The Head Clinical Care (HCC)
University Teaching Hospitals
Private Bag RW 1X
Ridgeway
Lusaka, Zambia, 10101

September 2, 2022

REF: REQUEST FOR ACCESS TO PATIENTS' OLD RADIOLOGY REPORTS AND IMAGES FOR THE PURPOSE OF SYSTEMATICALLY ORGANIZING THEM TO FACILITATE THEIR EFFICIENT STORAGE AND RETRIEVAL AS PART OF A PILOT RESEARCH PROJECT — ENTERPRISE MEDICAL IMAGING FOR STREAMLINED RADIOLOGICAL DIAGNOSIS IN ZAMBIAN PUBLIC HEALTH FACILITIES.

Dear Sir,

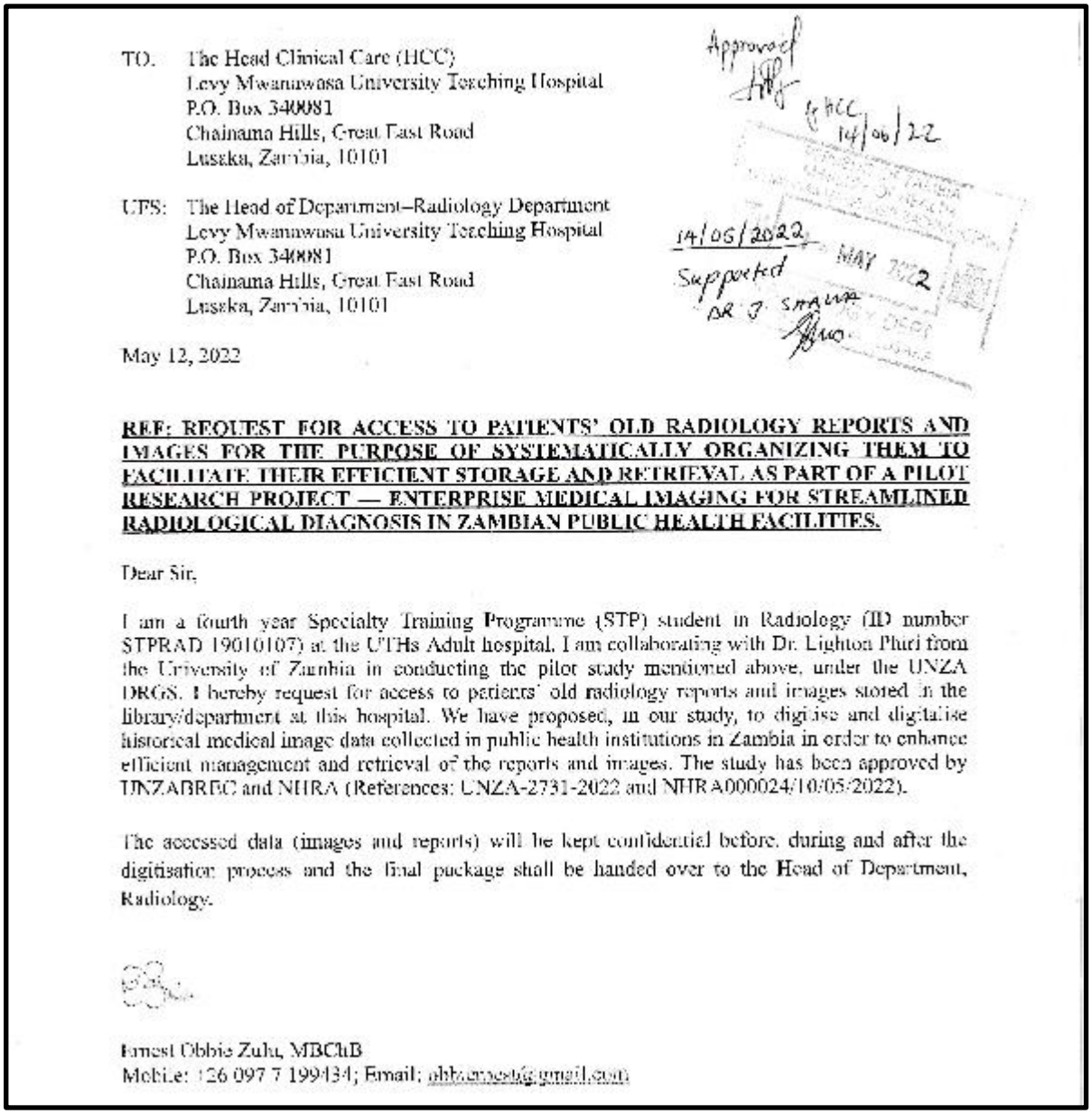
I am a fourth year Specialty Training Programme (STP) student in Radiology (ID number STPRAD 19010107) at the UTHs Adult hospital. I am collaborating with Dr. Lighton Phiri from the University of Zambia in conducting the pilot project mentioned above under the UNZA DRGS. We hereby request for access to patients' old radiology reports and images stored in the library/department at this hospital. The UTHs have been deliberately selected for piloting of this project, which was approved by UNZABREC and NHRA (References: UNZA-2731-2022 and NHRA000024/10/05/2022).

The accessed data (images and reports) will be kept confidential before, during and after the digitisation process and the final package shall be handed over to the Head of Department, Radiology.

I look forward to hearing from you.

Ernest Obbie Zulu, MBChB
Mobile: +26 097 7 199434; Email: obbiernest@gmail.com

Appendix K: Levy Mwanawasa Teaching Hospital Ethical Clearance Approval



TO. The Head Clinical Care (HCC)
Levy Mwanawasa University Teaching Hospital
P.O. Box 340081
Chainama Hills, Great East Road
Lusaka, Zambia, 10101

CFS: The Head of Department–Radiology Department
Levy Mwanawasa University Teaching Hospital
P.O. Box 340081
Chainama Hills, Great East Road
Lusaka, Zambia, 10101

Approved HCC 14/06/22

14/05/2022 MAY 2022 Supported DR J. SIMUMA

May 12, 2022

REF: REQUEST FOR ACCESS TO PATIENTS' OLD RADIOLOGY REPORTS AND IMAGES FOR THE PURPOSE OF SYSTEMATICALLY ORGANIZING THEM TO FACILITATE THEIR EFFICIENT STORAGE AND RETRIEVAL AS PART OF A PILOT RESEARCH PROJECT — ENTERPRISE MEDICAL IMAGING FOR STREAMLINED RADIOLOGICAL DIAGNOSIS IN ZAMBIAN PUBLIC HEALTH FACILITIES.

Dear Sir,

I am a fourth year Specialty Training Programme (STP) student in Radiology (ID number STPRAD 19010107) at the UTHs Adult hospital. I am collaborating with Dr. Lighton Phiri from the University of Zambia in conducting the pilot study mentioned above, under the UNZA DRGS. I hereby request for access to patients' old radiology reports and images stored in the library/department at this hospital. We have proposed, in our study, to digitise and digitalise historical medical image data collected in public health institutions in Zambia in order to enhance efficient management and retrieval of the reports and images. The study has been approved by UNZABREC and NHRA (References: UNZA-2731-2022 and NHRA000024/10/05/2022).

The accessed data (images and reports) will be kept confidential before, during and after the digitisation process and the final package shall be handed over to the Head of Department, Radiology.

Ernest Obbie Zulu, MBChB
Mobile: +26 097 7 199434; Email: obbiezulu@gmail.com

Appendix L: Clinicians' Awareness and Perceptions of the Use of Artificial Intelligence in Medical Imaging in Zambia (Biodata and Challenges Sections)

Institutional and Personal Demographic Details

The questions in this section of the questionnaire are aimed at eliciting demographic details about yourself and the institution/facility/hospital you work with/for.

Part 1: Facility Particulars

Facility Name *

Your answer

Facility Location (Province) *

Choose

Current Facility Level *

Choose

Facility Ownership *

- ○ GRZ
- ○ Military
- ○ NGO (e.g. Mission Hospital)
- ○ Private

Part 2: Participant Particulars

Gender *

- ○ Female
- ○ Male

Age Group *

- ○ 18-24
- ○ 25-30
- ○ 31-40
- ○ 41-50
- ○ 51-60
- ○ Above 60

Highest Level of Education *

- ◯ Certificate
- ◯ Diploma
- ◯ Advanced Diploma
- ◯ Bachelor's Degree
- ◯ Master's Degree
- ◯ PhD

Profession *

- ◯ Medical Imager (Radiographer, Radiography Technician, Sonographer, Radiation Physicist, etc))
- ◯ Clinical Officer/Medical Lincentiate
- ◯ Radiology Resident
- ◯ Radiologist
- ◯ Medical Doctor (All specialities)
- ◯ Other: _____

Years of Practice *

- ◯ Less Than 1 Year
- ◯ 2 - 5 Years
- ◯ 6 - 10 Years
- ◯ 11 - 20 Years
- ◯ More Than 20 Years

What previous Level Hospitals have you worked at (Select All Applicable)? *

- ☐ Level 1 Hospital
- ☐ Level 2 Hospital
- ☐ Level 3 Hospital
- ☐ Hospice
- ☐ Dental Clinic
- ☐ Diagnostic Centre
- ☐ Eye clinic
- ☐ Fertility clinic
- ☐ First-aid stations
- ☐ General clinic
- ☐ Health Centre
- ☐ Health Post
- ☐ Mobile clinic
- ☐ Optic Clinics
- ☐ Others
- ☐ Physiotherapy
- ☐ Rehabilitation centre
- ☐ Specimen Collection centre

Experience Interpreting Medical Images

Do you have experience interpreting medical images? *

- ◯ Yes
- ◯ No

Experience Interpreting Medical Images

What modality do you have experience in? *

- [ ] Plain X-ray Radiographs
- [ ] Mammographs
- [ ] Computerized Tomography (CT) scan
- [ ] Magnetic Resonance Imaging (MRI)
- [ ] Ultrasound
- [ ] Scintigraphy
- [ ] Hybrid (SPECT-CT, PET-CT, PET-MRI)
- [ ] Fluoroscopy
- [ ] Other: _____

Level of confidence in interpreting the modality ticked in previous question. *

- ◯ Very low
- ◯ Low
- ◯ Average
- ◯ High
- ◯ Very high

To what extent do you agree with the following statements? *

	Strongly agree	Somewhat agree	Neither agree or disagree	Disagree	Strongly disagree
The volumes of studies I sometimes interpret affects the quality of my reports.	○	○	○	○	○
I often have more imaging studies to interpret than I can reasonably handle in a workday.	○	○	○	○	○
I frequently experience time pressure when interpreting medical images.	○	○	○	○	○
My effectiveness is reduced when patient clinical history is incomplete or missing.	○	○	○	○	○
I am able to consistently accurately interpret medical images.	○	○	○	○	○
I am able to complete my reporting within the expected turnaround time.	○	○	○	○	○

General Awareness of Artificial Intelligence (AI) in Medical Imaging/Radiology

This section is aimed at eliciting information related to knowledge of the use of AI in Medical Imaging.

Part 3: Awareness of Artificial Intelligence

Have you ever heard of Artificial Intelligence? (e.g. ChatGPT, Gemini, Meta AI) *

- ◯ Yes
- ◯ No

To what extent are you aware of the use of artificial intelligence for the following radiological tasks? *

	Not at all aware	Slightly aware	Somewhat aware	Moderately aware	Extremely aware
Interpretation of medical images (e.g. X-rays)?	◯	◯	◯	◯	◯
Writing of radiological reports	◯	◯	◯	◯	◯

Have you ever used artificial intelligence to interpret medical images or generate reports?

- ◯ Yes
- ◯ No

Part 4: Artificial Intelligence Perceptions

To what extent to you agree with the following statements *

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Medical imaging/radiological diagnoses by Artificial Intelligence can be trusted	◯	◯	◯	◯	◯
Using Artificial Intelligence will result in loss of employment	◯	◯	◯	◯	◯
Using Artificial Intelligence could result in faster interpretation of reports	◯	◯	◯	◯	◯
Using Artificial Intelligence could potentially reduce the workload	◯	◯	◯	◯	◯
Using Artificial Intelligence could potentially result in accurate diagnoses	◯	◯	◯	◯	◯

Appendix M: Biodata and Consent form for Classification Tool Interaction

Section 1 of 6

Biodata Participant information and Consent Form

Test Study for the Semi-Automated Interpretation of Medical Images

Dear Respondent,
Thank you very much for taking the time to be part of this study.
I am a Masters in Computer Science student at the University of Zambia. This questionnaire is a data collection tool for a research paper "". The data collected will be used by the researcher to evaluate the usefulness and effectiveness of the method of Artificial Intelligence intervention for semi-automated interpretation of medical images.

This questionnaire is to be responded to by Radiologists, Radiology Registrars and other medical doctors working at the University Teaching Hospitals in Zambia.

This study has been cleared by the University of Zambia Biomedical Research Ethics Committee (UNZABREC) (approval code: UNZA-2731-2022) and authority has been granted by the National Health Research Authority (Ref. No.: NHRA000024/10/05/2022). Additionally, the study has been cleared by the Ministry of Health.

For further information and/or clarification, contact the following:

Investigator
Malaizyo Gabriel Muzumala
Graduate Student, University of Zambia
Email: malaizyo.muzumala@cs.unza.zm
Phone: +260972335056
WhatsApp: +260972335056

Section 2 of 6

Research Ethics Committee Clearance Certificate

Description (optional)

Image title



UNIVERSITY OF ZAMBIA
BIOMEDICAL RESEARCH ETHICS COMMITTEE

Telephone: +260 977925304
Telegrams: UNZA, LUSAKA
Telex: UNZALU ZA 44370
Fax: + 260-1-250753
Federal Assurance No. FWA00000338

Ridgeway Campus
P.O. Box 50110
Lusaka, Zambia
E-mail: unzarec@unza.zm
IRB00001131 of IORG0000774

5th May, 2022

Your REF. No. 2731-2022

Section 3 of 6

Grant of Authority to Conduct Research

Description (optional)

Image title



NATIONAL HEALTH RESEARCH AUTHORITY
Paediatric Centre of Excellence, University Teaching Hospital, P.O. Box 30075, LUSAKA
Chalala Office Lot No. 18961/M, Off Kasama Road, P.O. Box 30075, LUSAKA
Tell: +260211 250309 | Email: znhrasec@nhra.org.zm | www.nhra.org.zm

Ref No: NHRA000024/10/05/2022 **Date: 10th May, 2022**

Section 4 of 6

Participant Information Sheet

TITLE OF RESEARCH:

Test Study for the Semi-Automated Interpretation of Medical Images

BACKGROUND:

The Republic of Zambia faces a critical shortage of radiologists, and this scarcity highlights challenges in radiological diagnosis, forming the basis for a pilot project exploring the effectiveness of Artificial Interpretation techniques to assist in image interpretation. The study aims to address healthcare disparities by leveraging technology, focusing specifically on Artificial Intelligence techniques in chest x-ray Interpretation. The tool you will interact with will allow you to interact with different models and methods with the goal of investigating the perceived workload, usefulness integrating the underlying models streamlining interpretation, reducing time, improving accuracy, and enhancing workflow efficiency, ultimately contributing to superior patient care.

PURPOSE OF THE STUDY:

This study aims at evaluating the usefulness and ease of using the AI interventions for Semi-Automatic Interpretation of Medical Images at the University Teaching Hospitals in Zambia.

DESCRIPTION OF THE STUDY AND YOUR INVOLVEMENT:

The study test aims to find out the perceived usefulness and workload of the AI method. Participants will be handed an instruction sheet outlining the tasks to be performed and a questionnaire will be provided where each participant will be required to fill after having interacted with the system. The task execution and questionnaire filling are expected to last a maximum of 20 minutes.

CONFIDENTIALITY:

The researchers will guarantee confidentiality of the research findings and participants' identity.

VOLUNTARY PARTICIPATION AND WITHDRAWAL:

Consent will be obtained from all stakeholders or participants involved in the interview and observation process and participants can withdraw from the process if not comfortable.

RISKS AND BENEFITS

There are no known potential risks in taking part in this interview. There are no personal benefits associated with participation in this study. However, end-result benefits include solving challenges that radiologists are faced with during the interpretation of medical images and by so doing, help improve the radiological workflow.

CONTACTS FOR QUESTIONS (Names, addresses and phone numbers)

Supervisor
Dr Ernest Zulu
University Teaching Hospitals
Email: obbiernest@gmail.com
Phone: +260977199434
WhatsApp: +260977199434

Supervisor
Dr. Lighton Phiri
The University of Zambia
Email: lighton.phiri@unza.zm
Phone: +260961352662
WhatsApp: +260961352662

Section 5 of 6

Participant Consent Form

(Translated into vernacular if necessary)

REFERENCE TO PARTICIPANT INFORMATION SHEET:

1. Make sure that you read the Information Sheet carefully, or that it has been explained to you to your satisfaction.
2. Your participation in this research is entirely voluntary, i.e. you do not have to participate if you do not wish to.
3. Refusal to take part will involve no penalty or loss of services to which you are otherwise entitled.
4. If you decide to take part, you are still free to withdraw at any time without penalty or loss of services and without giving a reason for your withdrawal.
5. You may choose not to answer particular questions that are asked in the study. If there is anything that you would prefer not to discuss, please feel free to say so.
6. The information collected in this interview will be kept strictly confidential.
7. If you choose to participate in this research study, your signed consent is required below before I proceed with the interview with you.

VOLUNTARY CONSENT

I have read (or have had explained to me) the information about this research as contained in the Participant Information Sheet. I have had the opportunity to ask questions about it and any questions I have asked have been answered to my satisfaction.

I now consent voluntarily to be a participant in this project and understand that I have the right to end the interview at any time, and to choose not to answer particular questions that are asked in the study.

My signature below says that I am willing to participate in this research:

Do you consent to take part in this study? *

- ○ Yes
- ○ No

Participant's Name: *

Short answer text

Participant Number (On Instruction Sheet or given) *

Short answer text

Consent Date: *

Month, day, year

Section 6 of 6

Demographic Details

Description (optional)

Participant gender *

- Male
- Female

Profession *

- Radiologist
- Radiology Resident
- Medical Doctor (All specialties)

Years of practice? *

- Less than a year
- 1 to 5 years
- 5 to 10 years
- 10 to 15 years
- 15 to 20 years
- More than 20 years

Appendix N: University of Zambia Biomedical Research Ethics Committee Approval Renewal 2023.

UNIVERSITY OF ZAMBIA
BIOMEDICAL RESEARCH ETHICS COMMITTEE

Telephone: 260-1-256067
Telegrams: UNZA, LUSAKA
Telex: UNZALU ZA 44370
Fax: + 260-1-250753
Federal Assurance No. FWA00000338

Ridgeway Campus
P.O. Box 50110
Lusaka, Zambia
E-mail: unzarec@unza.zm
IRB00001131 of IORG0000774

11th May, 2023.

Ref. No. 2731-2022.

Dr. Ernest Obbie Zulu,
University Teaching Hospitals,
Adult Hospital Department of Radiology,
P/Bag RW 1X,
Ridgeway,
Lusaka.

Dear Dr. Zulu,

RE: APPLICATION FOR RENEWAL OF ETHICAL CLEARANCE APPROVAL: "ENTERPRISE MEDICAL IMAGING FOR STREAMLINED RADIOLOGICAL DIAGNOSIS IN ZAMBIAN PUBLIC HEALTH FACILITIES" (REF. NO. 2731-2022)

We acknowledge receipt of the request for renewal to the aforementioned study.

Renewal is hereby given for a period of one year from 5th May 2023 to 4th May 2024.

Yours sincerely,

Munsaka.

Sody Mweetwa Munsaka, BSc., MSc., PhD
CHAIRPERSON
Tel: +26099925304
E-Mail: s.munsaka@unza.zm

Appendix O: University of Zambia Biomedical Research Ethics Committee Approval Renewal 2025.

UNIVERSITY OF ZAMBIA
BIOMEDICAL RESEARCH ETHICS COMMITTEE

Telephone: 260-1-256067
Telegrams: UNZA, LUSAKA
Telex: UNZALU ZA 44370
Fax: + 260-1-250753
Federal Assurance No. FWA00000338

Ridgeway Campus
P.O. Box 50110
Lusaka, Zambia
E-mail: unzarec@unza.zm
IRB00001131 of IORG0000774

21st July, 2025.

Ref. No. 2731-2022.

Dr. Ernest Obbie Zulu,
University Teaching Hospitals,
Adult Hospital Department of Radiology,
P/Bag RW 1X,
Ridgeway,
Lusaka.

Dear Dr. Zulu,

RE: APPLICATION FOR RENEWAL OF ETHICAL CLEARANCE APPROVAL: "ENTERPRISE MEDICAL IMAGING FOR STREAMLINED RADIOLOGICAL DIAGNOSIS IN ZAMBIAN PUBLIC HEALTH FACILITIES" (REF. NO. 2731-2022)

We acknowledge receipt of the request for renewal to the aforementioned study.

Renewal is hereby given for a period of two years from:

i. 5th May 2024 to 4th May 2025 and from
ii. 5th May 2025 to 4th may 2026

Yours sincerely,

Munsaka.

Prof. Sody Mweetwa Munsaka, BSc., MSc., PhD
CHAIRPERSON
Tel: +26099925304
E-Mail: s.munsaka@unza.zm