MedScan

Multimodal Medical ReportAnalyzer

PROJECT REPORT

By

KAPIL (250870828005)

Kamalesh Mukherjee(250870828006)

Submitted in partial fulfilment of the requirements for the award of

PG DIPLOMA IN ARTIFICIAL INTELLIGENCE (PG-DAI)

C-DAC ACTS
RESEARCH PARK, IIT GUWAHATI CAMPUS
ASSAM

CERTIFICATE

This is to certify that the project report titled **"MedScan"** submitted by **Kapil (250870828005), Kamalesh Mukherjee(250870828006)** in partial fulfilment of the requirements for the award of the **PG DIPLOMA IN ARTIFICIAL INTELLIGENCE (PG-DAI)** is a Bonafide record of work carried out under my supervision and guidance.

Mr. David Ray **Ms. Ruchika Nath**

ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to our project guide **Mr. David Ray and Ms. Ruchika Nath**, for their invaluable guidance and constant support throughout the duration of this project.

We would also like to extend our thanks to **Mr. Alok Dey**, Head of Department, for providing us with the opportunity and the necessary facilities to undertake this project. We are also grateful to **C-DAC CINE Guwahati** for fostering an academic environment conducive to learning and innovation.

Lastly, We are thankful to our friends and family for their unwavering encouragement and motivation.

Date: Jan 2026

Place: C-DAC ACTS Guwahati

Kapil (250870828005)

Kamalesh Mukherjee(250870828006)

ABSTRACT

In recent years, the healthcare sector has witnessed a rapid increase in the use of digital medical records, diagnostic reports, and imaging data. Hospitals and diagnostic centers routinely generate large volumes of medical reports in the form of PDFs, which may include scanned documents, structured tables, unstructured clinical text, and embedded medical images such as X-rays, CT scans, MRI scans, ultrasound images, and ECG reports. Interpreting these reports manually is time-consuming, error-prone, and often challenging for patients who lack medical expertise.

The **Multimodal Medical Report Analyzer** proposed in this project aims to address these challenges by providing an end-to-end automated system for analyzing medical PDF reports. The system is designed to process multiple data modalities present in a single report, including textual content, tabular laboratory results, and medical images. It employs a combination of direct PDF text extraction and Optical Character Recognition (OCR) techniques to handle both digital and scanned reports. Extracted tables are further processed to identify laboratory data, clean inconsistencies, and flag abnormal values based on reference ranges.

To structure unorganized medical text, the system utilizes a **hybrid rule-based Natural Language Processing (NLP) approach**, which performs sentence-level intent detection and organizes information into a unified, structured report format. In parallel, embedded medical images are extracted from PDF files and classified using Convolutional Neural Network (CNN)–based models to identify the image modality, such as X-ray, CT, MRI, ultrasound, ECG, or non-medical images.

To enhance usability and understanding, the structured report is passed through a controlled **Large Language Model (LLM)–based explanation layer**, which generates patient-friendly and clinician-oriented summaries while strictly enforcing medical safety constraints. The system is deployed through a user-friendly Streamlit web interface, enabling users to upload reports, view extracted data, analyze results, and obtain simplified explanations in real time.

The proposed system demonstrates how multimodal data processing, rule-based NLP, computer vision, and controlled language models can be effectively integrated into a single framework to support medical report interpretation. This project provides a scalable, explainable, and safety-aware solution that can assist patients, clinicians, and healthcare institutions in understanding complex medical reports more efficiently.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Background of the Study

The healthcare industry has undergone a significant digital transformation in recent years, with hospitals and diagnostic centers increasingly relying on electronic medical records and digitally generated diagnostic reports. Medical reports are commonly issued in Portable Document Format (PDF), as it allows easy sharing, storage, and standardization across systems. These reports often contain a combination of unstructured clinical text, structured laboratory tables, and embedded diagnostic images such as X-rays, CT scans, MRI scans, ultrasound images, and ECG recordings.

While digital reports have improved accessibility, their interpretation remains a complex and time-consuming task. Medical reports are usually reviewed manually by healthcare professionals, requiring careful reading and domain-specific expertise. Moreover, different types of reports demand interpretation by different specialist doctors. For example, laboratory reports are analyzed by pathologists, imaging reports by radiologists, and ECG reports by cardiologists. This specialization-based workflow increases dependency on multiple experts and adds to the overall workload in healthcare systems.

With the growing volume of diagnostic data, there is an increasing need for intelligent systems that can assist in organizing, analyzing, and summarizing medical reports. Such systems can support healthcare professionals by reducing repetitive manual work and can also help patients better understand their medical reports before consulting a doctor.

1.2 Problem Statement

Modern medical reports are heterogeneous in nature, often combining scanned pages, digitally generated text, structured tables, and medical images within a single document. Extracting meaningful information from such reports requires significant manual effort and careful attention to detail. This process becomes more challenging when reports originate from different diagnostic domains, each requiring interpretation by specialized medical professionals.

The reliance on multiple specialists for reading and interpreting different sections of medical reports can lead to delays, increased workload, and inefficiencies, particularly in high-volume clinical environments. Patients, on the other hand, may find medical reports difficult to understand due to complex terminology and fragmented presentation of information.

Additionally, existing automated systems often focus on only one type of data, such as text or images, and fail to handle the multimodal nature of real-world medical reports. There is a lack of unified solutions capable of processing text, tables, and images together while presenting the information in a structured and easy-to-understand manner.

Therefore, there is a need for an assistive system that can automatically analyze multimodal medical reports, organize extracted information coherently, and provide simplified explanations, thereby reducing manual workload and improving accessibility without replacing the role of medical professionals.

1.3 Objectives of the Project

The primary objectives of the Multimodal Medical Report Analyzer are as follows:

- To develop an automated system capable of analyzing medical PDF reports containing text, tables, and images.
- To extract textual content from both digital and scanned medical reports using appropriate extraction and OCR techniques.
- To identify and process laboratory tables, clean extracted data, and flag abnormal values based on reference ranges.
- To organize unstructured medical text into a structured format using a rule-based Natural Language Processing approach.
- To extract and classify embedded medical images into categories such as X-ray, CT, MRI, ultrasound, ECG, or non-medical images.
- To generate patient-friendly and clinician-oriented summaries using a controlled language model with safety constraints.
- To provide a user-friendly web interface for uploading reports and viewing analyzed results.

1.4 Scope of the Project

The scope of this project is focused on assisting in the preliminary analysis and understanding of medical reports across multiple diagnostic domains. The system supports medical PDF reports that include laboratory investigations, clinical narratives, and embedded diagnostic images, which traditionally require interpretation by different specialist doctors.

By integrating text extraction, table analysis, image classification, and structured report generation into a single framework, the system aims to reduce repetitive manual effort involved in reading and organizing medical reports. The application is designed to support healthcare professionals, general practitioners, and patients by presenting medical information in a clear and structured manner.

The proposed system functions as an assistive and decision-support tool. It does not provide medical diagnoses or treatment recommendations and is not intended to replace healthcare professionals. Instead, it enhances efficiency by streamlining report interpretation and improving accessibility of medical information.

1.5 Significance of the Study

The Multimodal Medical Report Analyzer holds significance in addressing practical challenges faced in modern healthcare environments. By reducing manual effort required to interpret complex medical reports, the system can help optimize the time and workload of medical professionals. It also empowers patients by providing simplified explanations that improve understanding of their medical reports.

From a technical perspective, the project demonstrates the effective integration of rule-based Natural Language Processing, computer vision techniques, and controlled language models within a real-world application. The system emphasizes explainability, safety, and modular design, making it suitable for future enhancements and institutional deployment.

This project also serves as a practical implementation of advanced computing concepts applied to a critical real-world domain, providing valuable insights into building reliable, scalable, and responsible medical software systems.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The rapid digitization of healthcare systems has led to a significant increase in the generation and storage of electronic medical records and diagnostic reports. Medical reports today often exist in digital formats such as PDFs and include diverse forms of data, including unstructured clinical text, structured laboratory tables, and diagnostic images. The literature review explores existing research and systems related to medical report analysis, document processing, Natural Language Processing (NLP), computer vision in healthcare, and automated medical summarization.

The purpose of this chapter is to examine previous work in these areas, identify existing approaches, and highlight the limitations that motivate the development of the proposed Multimodal Medical Report Analyzer.

2.2 Medical Report Digitization and Analysis

Medical report digitization has been a major focus of healthcare informatics research. Early systems primarily relied on electronic health record (EHR) platforms to store structured data entered manually by healthcare professionals. However, many diagnostic reports are still generated as scanned documents or semi-structured PDFs, making automated processing challenging.

Several studies have emphasized the need for automated tools capable of extracting information from medical documents. Techniques such as Optical Character Recognition (OCR) have been widely adopted to convert scanned reports into machine-readable text. While OCR systems have improved significantly, medical documents often contain complex layouts, tables, and varying fonts, which can lead to extraction errors. As a result, post-processing and validation mechanisms are commonly required to improve accuracy.

2.3 Natural Language Processing in Healthcare

Natural Language Processing plays a crucial role in extracting meaningful insights from unstructured medical text. NLP techniques have been used for tasks such as clinical entity recognition, report classification, and summarization. Rule-based NLP systems were among the earliest approaches used in healthcare due to their transparency, reliability, and ease of validation.

More recent studies have explored machine learning and deep learning–based NLP models to analyze clinical narratives. While such models can achieve high accuracy, they often require large annotated datasets and may lack explainability, which is a critical requirement in medical applications. As a result, many healthcare systems still rely on hybrid approaches that combine rule-based NLP with statistical or learning-based methods to balance accuracy and interpretability.

2.4 Analysis of Laboratory Reports

Laboratory reports are a fundamental component of medical diagnostics and typically consist of structured tables containing test names, measured values, units, reference ranges, and abnormality flags. Research in this area has focused on extracting tabular data from medical documents and mapping results to standardized medical terminologies.

Several tools and frameworks have been proposed to detect and extract tables from PDFs. However, scanned reports and poorly formatted documents remain challenging. Studies have highlighted the importance of data cleaning and normalization to ensure consistency across laboratory reports. Automated flagging of abnormal values has also been explored as a means of assisting clinicians in quickly identifying critical results.

2.5 Computer Vision in Medical Imaging

Medical imaging plays a vital role in diagnosis and treatment planning. Computer vision techniques, particularly Convolutional Neural Networks (CNNs), have been widely used for tasks such as image classification, segmentation, and anomaly detection in medical images. Applications include classification of X-rays, CT scans, MRI scans, ultrasound images, and ECG plots.

Most existing research focuses on disease detection and diagnostic prediction. However, fewer studies address the problem of automatically identifying the type of medical image embedded within medical documents. Image classification at the modality level can serve as an important preprocessing step for organizing reports and routing them to appropriate specialists. The use of CNN-based classifiers for modality detection provides a reliable and scalable solution for handling large volumes of medical images.

2.6 Multimodal Medical Document Processing

Recent research has increasingly focused on multimodal approaches that integrate text, tables, and images within a single analytical framework. Multimodal medical document processing aims to capture the full context of medical reports rather than analyzing each component independently.

Existing multimodal systems often suffer from limitations such as high computational complexity, lack of interpretability, or reliance on fully trained end-to-end models. Studies suggest that modular architectures, where each modality is processed independently and later integrated, offer better flexibility and maintainability. Hybrid pipelines that combine deterministic rules with machine learning components have been shown to be effective in real-world healthcare environments.

2.7 Automated Medical Report Summarization

Automated summarization of medical reports has gained attention as a means of improving accessibility for patients and reducing cognitive load for clinicians. Early summarization systems were rule-based and focused on extracting key sentences from reports. More recent approaches employ Large Language Models (LLMs) to generate natural language summaries.

Despite their capabilities, LLM-based systems pose challenges related to factual accuracy, hallucination, and safety. Literature emphasizes the importance of controlled generation, domain constraints, and explicit disclaimers when using language models in healthcare. Systems that restrict LLM usage to explanation and summarization tasks, rather than diagnosis, are considered safer and more acceptable in clinical settings.

2.8 Research Gap

several gaps remain. Many existing systems focus on a single data modality, such as text or images, and fail to address the heterogeneous nature of real-world medical reports. Systems that rely entirely on machine learning models often lack transparency and are difficult to validate in medical environments.

There is a need for an integrated, multimodal system that can process text, tables, and images together while maintaining explainability and safety. Additionally, there is limited work on systems that assist in reducing manual workload across multiple medical specializations without attempting to replace medical professionals.

The proposed Multimodal Medical Report Analyzer addresses these gaps by adopting a hybrid, modular approach that combines rule-based NLP, computer vision, and controlled language models within a unified framework.

2.9 Summary

This chapter reviewed existing literature related to medical report digitization, NLP in healthcare, laboratory report analysis, medical image processing, multimodal document analysis, and automated summarization. The review highlights the limitations of current approaches and underscores the need for a comprehensive, explainable, and safety-aware system capable of analyzing complex medical reports. These findings form the foundation for the methodology and system design discussed in the subsequent chapters.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

Research methodology defines the systematic approach adopted to analyze the problem, design the solution, and implement the proposed system. It outlines the methods, tools, and techniques used to ensure that the project objectives are achieved in an efficient and reliable manner.

For the Multimodal Medical Report Analyzer, the research methodology focuses on understanding real-world medical report structures, identifying challenges in manual report interpretation, and designing a system capable of processing heterogeneous data modalities such as text, tables, and medical images. The methodology emphasizes accuracy, explainability, and safety, which are critical requirements in healthcare-related applications.

3.2 Research Approach

The project follows an **applied and descriptive research approach**. The applied aspect addresses a real-world healthcare problem by proposing an automated system for medical report analysis. The descriptive aspect involves studying existing medical report formats, diagnostic workflows, and current document-processing techniques to design an effective solution.

Rather than focusing on theoretical model development, the research emphasizes system-level integration of proven techniques such as Optical Character Recognition (OCR), rule-based Natural Language Processing, and computer vision methods. This approach ensures that the solution remains practical, interpretable, and suitable for real-world deployment.

3.3 Data Collection and Analysis

The data used in this project consists of medical PDF reports containing textual descriptions, laboratory tables, and embedded medical images. These reports represent a variety of diagnostic domains, including laboratory investigations and imaging-based diagnostics.

Data analysis involved examining common report layouts, identifying recurring patterns in laboratory tables, and understanding how medical images are embedded within documents. This analysis helped define extraction strategies and informed the design of preprocessing steps required to handle scanned documents, inconsistent formatting, and mixed data types.

No patient-identifying data was stored or analyzed beyond what was required for structural processing, ensuring ethical handling of medical information.

3.4 System Development Methodology

The project follows a **modular and incremental development approach**, inspired by the principles of the Waterfall Software Development Life Cycle (SDLC). Each phase was executed in a structured manner, ensuring clarity and traceability throughout development.

The major phases include:

1. **Requirement Analysis** – Identifying system objectives, functional requirements, and constraints related to medical data processing.

2. **System Design** – Designing the overall architecture, defining processing pipelines for text, tables, and images, and planning system integration.

3. **Implementation** – Developing individual modules for extraction, analysis, classification, and summarization.

4. **Testing** – Validating the correctness and reliability of each module and the overall system workflow.

5. **Evaluation** – Assessing system performance in terms of usability, accuracy of extraction, and effectiveness of report organization.

3.5 Tools and Technologies Used

The following tools and technologies were used in the development of the proposed system:

- **Programming Language:** Python
- **Web Framework:** Streamlit (for user interface and interaction)
- **PDF Processing:** PyMuPDF, pdf2image
- **Optical Character Recognition:** Tesseract OCR
- **Natural Language Processing:** Rule-based and heuristic NLP techniques
- **Table Extraction:** Camelot and OCR-based table parsing

- **Computer Vision:** Convolutional Neural Networks for image classification
- **Language Model Integration:** Controlled LLM for report explanation and summarization
- **Libraries and Frameworks:** PyTorch, OpenCV, Pandas, NumPy

These technologies were selected based on their reliability, community support, and suitability for medical document analysis.

3.6 Ethical and Safety Considerations

Given the sensitive nature of medical data, ethical and safety considerations were integral to the research methodology. The system was designed to function strictly as an assistive tool and does not perform diagnosis or treatment recommendations.

The language model component operates under strict constraints to prevent medical hallucinations, unsupported claims, or clinical advice. Clear disclaimers are included to emphasize that final interpretation of reports must be performed by qualified healthcare professionals.

Additionally, the system processes reports locally and does not require external data sharing, supporting data privacy and confidentiality.

3.7 Conclusion

This chapter outlined the research methodology adopted for the Multimodal Medical Report Analyzer. By combining an applied research approach with a structured development methodology, the project ensures practical relevance, safety, and explainability. The chosen tools and techniques enable effective handling of multimodal medical reports while maintaining ethical standards. The methodology provides a strong foundation for the system design and implementation discussed in the subsequent chapters.

CHAPTER 4: SYSTEM DESIGN

4.1 Introduction

System design represents a crucial phase in the development of any software application, as it translates conceptual requirements into a structured and functional architecture. A well-designed system ensures modularity, scalability, maintainability, and ease of integration of individual components.

In the Multimodal Medical Report Analyzer, system design focuses on handling heterogeneous medical data efficiently. The system is designed to process medical PDF reports containing unstructured text, structured tables, and embedded medical images, and to integrate the outputs of multiple processing modules into a unified and user-friendly interface.

4.2 Overall System Architecture

The proposed system follows a **modular and pipeline-based architecture**, where each module is responsible for handling a specific data modality or processing task. This separation of concerns allows independent development, testing, and enhancement of individual components without affecting the overall system.

At a high level, the system consists of the following layers:

- **Presentation Layer:** A web-based user interface built using Streamlit, allowing users to upload medical reports and view analysis results.
- **Processing Layer:** Core modules responsible for text extraction, table processing, image extraction, and image classification.
- **Analysis Layer:** Hybrid NLP-based report parsing and structured data generation.
- **Explanation Layer:** A controlled language model used for generating patient-friendly and clinician-oriented summaries.
- **Output Layer:** Visualization and presentation of extracted data, tables, images, and summaries.

4.3 System Workflow

The complete workflow of the proposed system is illustrated in **Figure 4.1**. The workflow demonstrates how a medical PDF report flows through various processing stages, from initial upload to final result presentation.

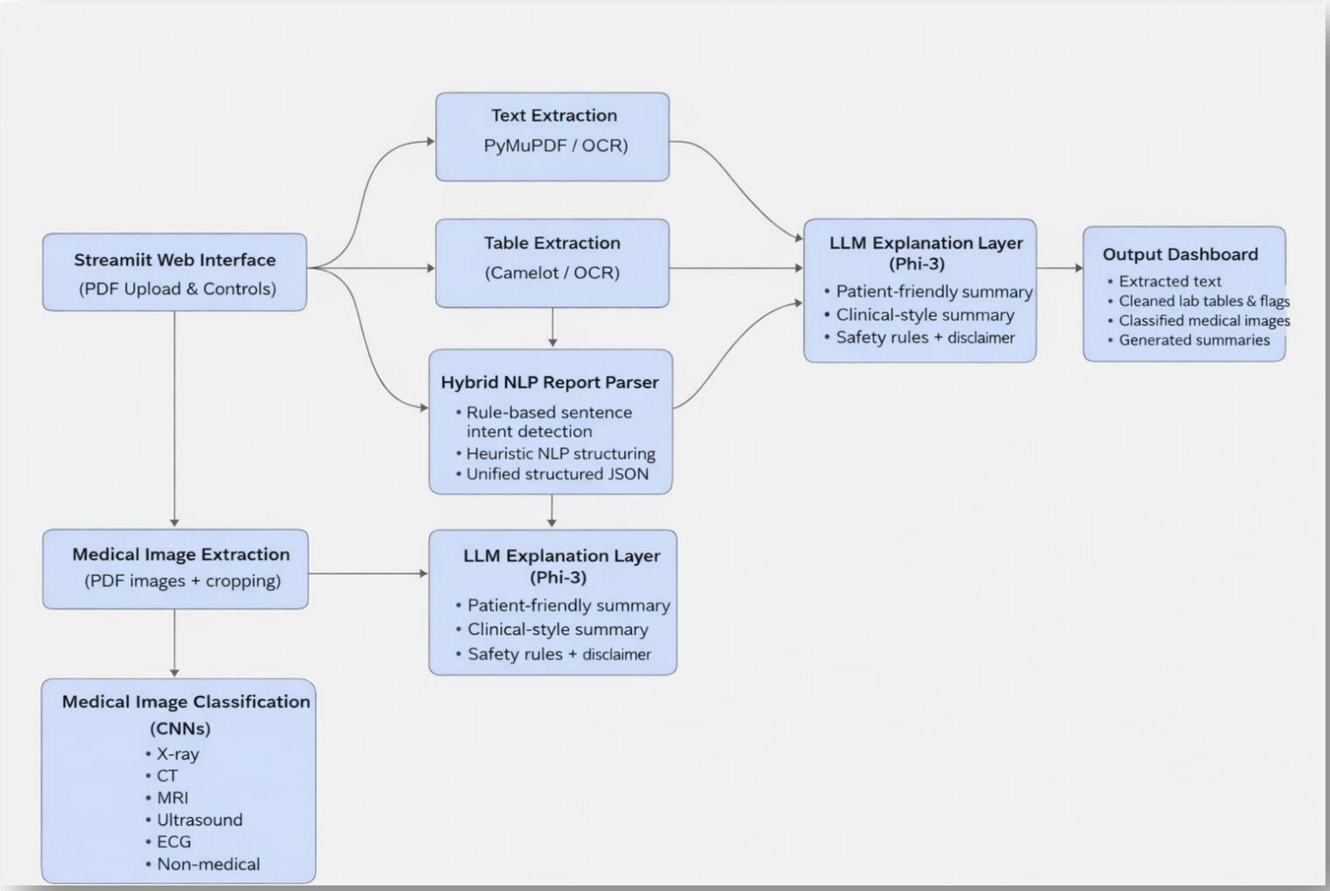


Figure 4.1: Overall Workflow of the Multimodal Medical Report Analyzer

As shown in Figure 4.1, the system begins with the user uploading a medical PDF report through the web interface. The report is then processed through multiple parallel pipelines based on data type:

Text Processing Pipeline:
The system first determines whether the PDF contains digitally extractable text or scanned content. Direct text extraction is applied where possible, while OCR is used for scanned pages. The extracted text is cleaned and normalized for further processing.

Table Processing Pipeline:
Tables are extracted from the PDF using table detection techniques. If the report contains laboratory tables, the system identifies the table type, cleans the extracted data, and flags abnormal values based on reference ranges.

Image Processing Pipeline:
Embedded images are extracted from the PDF. Medical images are isolated from report pages and passed to an image classification module that identifies the image modality, such as X-ray, CT, MRI, ultrasound, ECG, or non-medical images.
The outputs of these pipelines are integrated into a structured internal representation, which is then passed to the explanation layer for summary generation and final visualization.

4.4 Text Extraction and Preprocessing Design

Medical reports often exist in both digital and scanned formats. To handle this variability, the system uses a hybrid text extraction strategy. For digitally generated PDFs, direct text extraction techniques are applied. For scanned documents, Optical Character Recognition (OCR) is employed to convert images into machine-readable text.

Preprocessing steps such as noise removal, whitespace normalization, and page-wise organization are applied to ensure consistency across reports. This design ensures robust handling of diverse document formats commonly encountered in real-world healthcare settings.

4.5 Table Extraction and Laboratory Data Processing

Laboratory reports typically include structured tables containing test names, values, units, and reference ranges. The system design incorporates a dedicated table extraction module to detect and extract tabular data from medical PDFs.

Extracted tables undergo a cleaning and normalization process to remove inconsistencies and irrelevant rows. For laboratory tables, numerical values are compared against reference ranges to identify abnormal results. This structured approach allows the system to highlight important findings while maintaining transparency and explainability.

4.6 Image Extraction and Classification Design

Medical images embedded within reports are extracted using document-level image processing techniques. The system differentiates between full-page report images and diagnostic images to ensure accurate classification.

A CNN-based image classification module is used to identify the modality of each extracted image. The classification focuses on recognizing the type of medical image rather than diagnosing medical conditions. This design enables effective organization of reports and supports faster routing of information to relevant medical specialists.

4.7 Hybrid NLP-Based Report Parsing

To organize unstructured medical text, the system employs a **hybrid rule-based Natural Language Processing approach**. Sentence-level intent detection is used to categorize text into sections such as patient information, clinical history, findings, diagnosis, and recommendations.

This rule-based design was chosen to ensure interpretability, reliability, and safety, which are essential in healthcare applications. The parsed information is combined with structured table data and image classification results to form a unified internal report representation.

4.8 Explanation and Summary Generation Layer

The structured report is passed to a controlled language model that generates summaries tailored to different audiences. The system supports both patient-friendly explanations and clinician-oriented summaries.

Strict constraints are enforced to prevent the generation of diagnoses, treatment recommendations, or unsupported medical claims. Disclaimers are included to emphasize that the system functions as an assistive tool and does not replace professional medical judgment.

4.9 User Interface Design

The user interface is designed to be simple, intuitive, and accessible. Users can upload medical reports, preview PDF content, view extracted text and tables, inspect classified medical images, and read generated summaries through a single dashboard.

The interface presents results in a clear and organized manner, allowing users to interact with different components of the analyzed report without technical complexity.

4.10 Conclusion

This chapter presented the system design of the Multimodal Medical Report Analyzer, detailing its architecture, workflow, and individual modules. The modular and hybrid design enables effective processing of multimodal medical reports while maintaining safety, explainability, and scalability. The system design forms the foundation for the implementation details discussed in the next chapter.

CHAPTER 5: IMPLEMENTATION

5.1 Introduction

This chapter presents the implementation of the Multimodal Medical Report Analyzer through actual system outputs generated at different processing stages. Instead of reiterating the system design, this chapter demonstrates how each module performs in practice by showcasing intermediate and final results. The implementation validates the effectiveness of the proposed architecture and highlights the real-world applicability of the system.

5.2 Implementation Strategy

The system was implemented using a modular approach, where each functional unit was developed as an independent component. This strategy allows individual modules to be tested, debugged, and enhanced without affecting other parts of the system.

The major implementation stages include:

- User interface development for report upload and result visualization
- Text extraction and OCR integration
- Table extraction and laboratory data processing
- Medical image extraction and classification
- Hybrid NLP-based report parsing
- Controlled explanation and summarization

Each stage was integrated incrementally to ensure smooth end-to-end execution of the system.

5.3 Text Extraction and OCR Module Implementation

The system supports both digitally generated and scanned medical PDF reports. During implementation, direct text extraction was applied to digital PDFs, while Optical Character Recognition (OCR) was automatically triggered for scanned pages based on text availability thresholds.

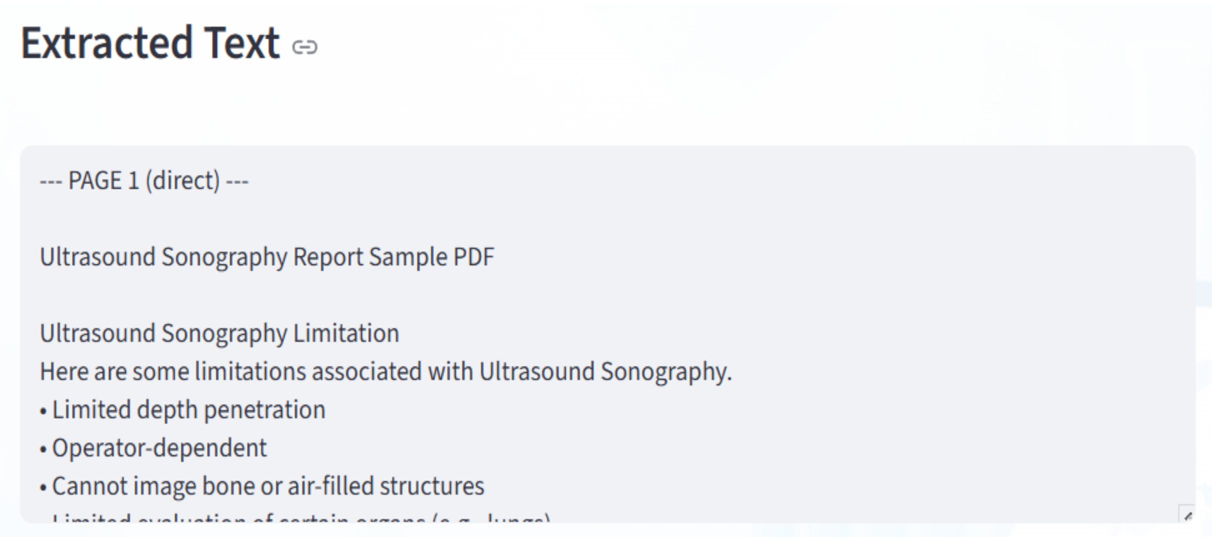


Figure 5.1: Extracted textual content from an uploaded medical PDF report

The extracted text is displayed within the user interface, preserving page-wise structure and content continuity. This output serves as the primary input for subsequent parsing and analysis stages.

5.4 Table Extraction and Laboratory Data Processing Implementation

Laboratory tables embedded within medical reports were extracted using a combination of table detection and OCR-based parsing techniques. The extracted data was further cleaned and normalized to remove inconsistencies and irrelevant entries.

Detected Table

	test	result	status	ref	unit
0	Total RBC count	5.00	Normal	None	None
1	Packed Cell Volume (PCV)	45	Normal	40-50	%
2	Mean Corpuscular Volume (MCV)	100	Normal	83-101	fl
3	es	30	Normal	27-32	pg
4	MCHC	33.00	Normal	None	None
5	Total WBC count	10000	Normal	4000 - 11000	cumm
6	Neutrophils	60	Normal	50-62	%
7	Lymphocytes	30	Normal	20-40	%
8	Eosinophils	2	Normal	00 - 06	%
9	eect cs	8	Normal	00-10	%

Figure 5.2: Laboratory table extraction and abnormal value flagging

As shown in Figure 5.2, the system highlights abnormal laboratory values using visual indicators, enabling quick identification of critical test results. This structured representation reduces manual effort in scanning complex laboratory reports.

5.5 Hybrid NLP-Based Report Parsing Results

To organize unstructured clinical text, the system applies a hybrid rule-based Natural Language Processing approach. The implementation classifies sentences based on intent and generates a structured representation of the report.

Figure 5.3: Structured medical report generated using hybrid NLP parsing

The structured output separates patient information, clinical history, findings, and recommendations, improving clarity and readability. This step forms the unified internal representation used across the system.

5.6 Medical Image Extraction and Classification Results

Medical images embedded within PDF reports were extracted and processed independently. The system isolates diagnostic images and classifies them according to modality using a CNN-based classification model.

Image Extraction & Classification

	image	type	confidence
0	uploaded_temp_page1_img1_crop1.png	X-ray	68.9
1	uploaded_temp_page1_img1_crop2.png	X-ray	95.1
2	uploaded_temp_page1_img1_crop3.png	X-ray	95

X-ray

X-ray (68.9%)

X-ray (95.1%)

X-ray (95.0%)

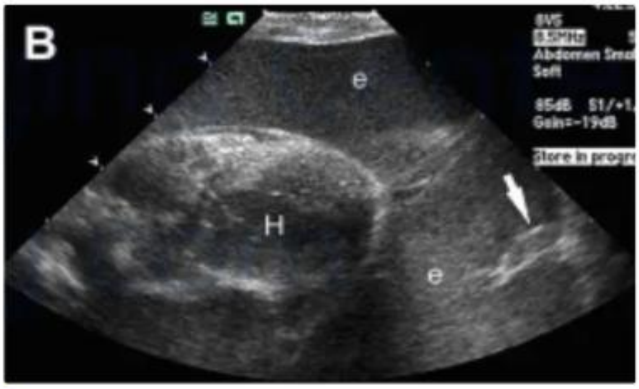
Ultrasound

The use_column_width parameter has been deprecated and will be removed in a future release. Please utilize the use_container_width parameter instead.

Ultrasound Confidence: 57.2%

› Model Breakdown

The use_column_width parameter has been deprecated and will be removed in a future release. Please utilize the use_container_width parameter instead.



Ultrasound Confidence: 83.0%

› Model Breakdown

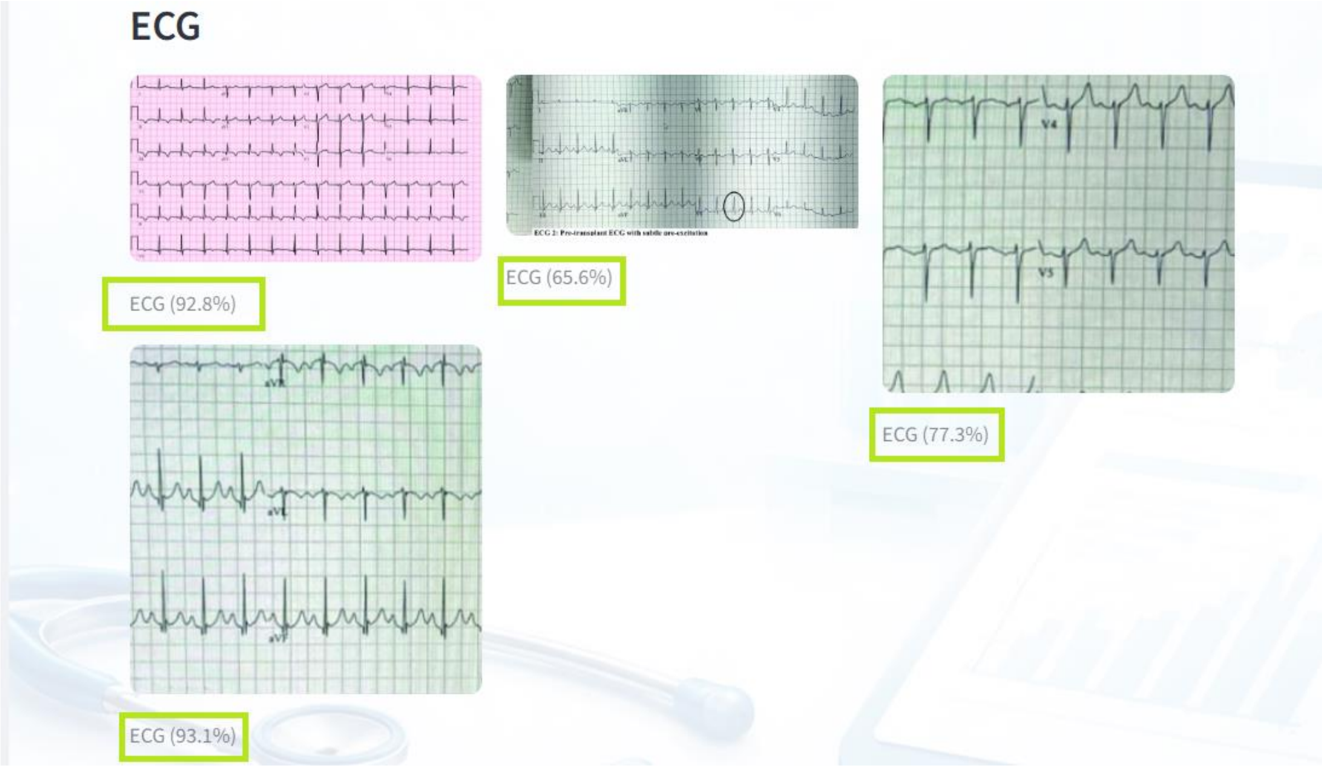
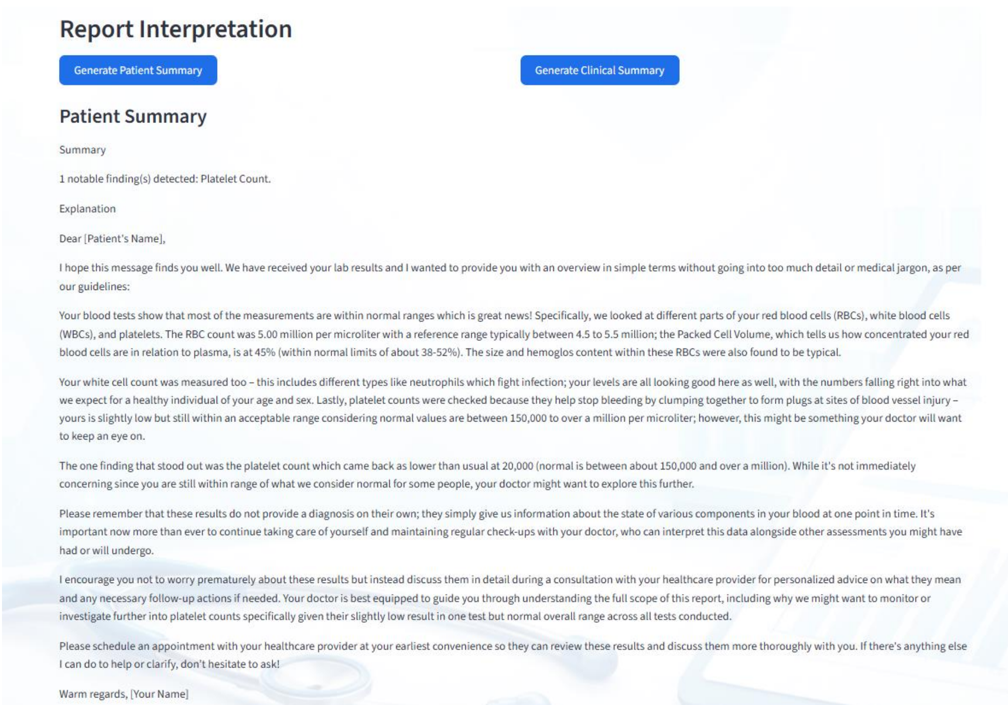


Figure 5.4: Extracted medical images with modality classification

Each image is labeled with its identified modality, such as X-ray, CT, MRI, ultrasound, or ECG. The classification focuses on image type recognition and does not attempt medical diagnosis.

5.6 Explanation and Summary Generation Results

The final structured report is passed to a controlled language model to generate user-friendly explanations. Two types of summaries are supported: patient-oriented and clinician-oriented.



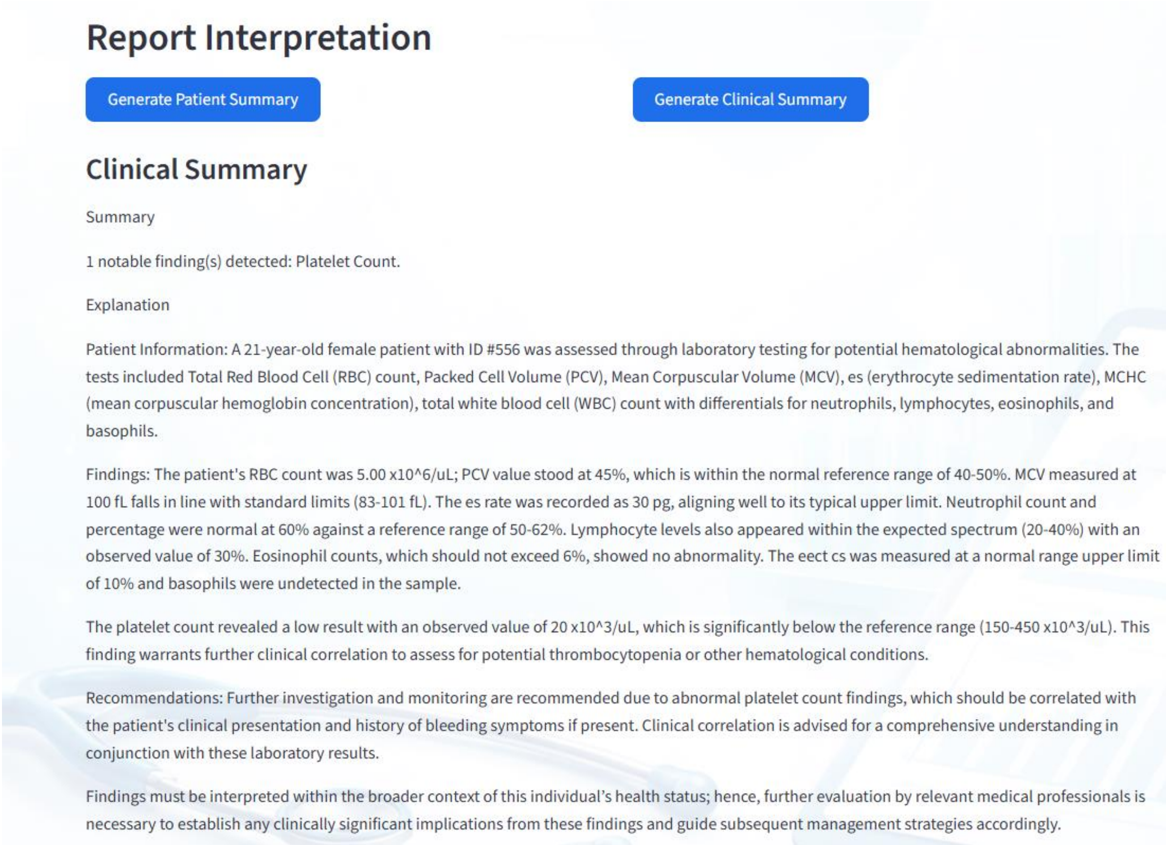


Figure 5.5: Patient-friendly and Clinical summary generated with safety constraints

The generated summaries adhere to strict safety rules and include disclaimers indicating that the output is for informational purposes only. This implementation enhances accessibility while maintaining medical responsibility.

5.7 System Integration and Error Handling

All individual modules were integrated into a single pipeline, ensuring smooth data flow from report upload to result generation. The system handles unsupported files, extraction failures, and unexpected inputs gracefully by displaying informative messages to the user while maintaining internal logs for debugging.

5.8 Conclusion

This chapter demonstrated the implementation of the Multimodal Medical ReportAnalyzer through practical system outputs. By presenting intermediate and final results at each processing layer, the implementation validates the effectiveness of the proposed system and highlights its applicability in real-world healthcare scenarios.

CHAPTER 6: TESTING AND VALIDATION

6.1 Introduction

Testing and validation are essential to ensure that the developed system functions correctly, reliably, and as intended under real-world conditions. This chapter describes the testing approach adopted for the Multimodal Medical Report Analyzer, focusing on functional correctness, module-level validation, and end-to-end system behavior.

Unlike traditional software systems, the proposed solution integrates multiple processing layers such as text extraction, table analysis, image classification, and report summarization. Therefore, testing was conducted in a progressive and modular manner to validate each component independently before full system integration.

6.2 Testing Strategy

The testing strategy followed a **step-by-step and incremental approach**. During the initial development phase, individual system components were implemented and tested within interactive `.ipynb` notebooks. This allowed rapid experimentation, debugging, and validation of logic for each module in isolation.

Once the functionality of a module was verified and stabilized, the corresponding code was migrated into structured `.py` files and integrated into the main application pipeline. This approach ensured that only tested and validated components were used in the final system.

Testing was primarily **functional and qualitative**, focusing on correctness of outputs, robustness across different report formats, and safe handling of edge cases.

6.3 Module-wise Functional Testing

Each system module was tested independently to ensure correct operation before integration.

- **Text Extraction Module:**
Tested on both digitally generated and scanned medical PDF reports to verify correct selection between direct text extraction and OCR-based extraction.

- **Table Extraction and Laboratory Processing Module:**
Tested on reports containing laboratory tables with varying layouts to confirm accurate extraction, data cleaning, and abnormal value flagging.

- **Hybrid NLP Parsing Module:**
Tested using multiple medical reports to validate sentence-level intent detection and correct structuring of extracted clinical information.

- **Medical Image Extraction and Classification Module:**
Tested on reports containing different types of embedded medical images to verify correct extraction and modality classification.

- **Explanation and Summarization Module:**
Tested to ensure summaries were generated safely, without diagnosis or treatment recommendations, and included appropriate disclaimers.

6.4 System-Level Testing on Multiple Medical Reports

After individual module validation, the complete system was tested end-to-end using **multiple real-world medical reports** representing different formats and diagnostic domains. These tests verified seamless interaction between modules and ensured correct data flow from report upload to final output generation.

The system was observed to handle variations in report structure, content density, and embedded data types effectively. Error conditions such as missing tables, absence of images, or low-quality scanned text were handled gracefully without system failure.

6.5 Result Summary

The testing process confirmed that the Multimodal Medical Report Analyzer functions reliably across diverse medical report formats. Module-wise testing ensured correctness at each processing layer, while end-to-end testing validated system integration and stability.

The incremental development and testing strategy, starting from notebook-based experimentation to full application integration, significantly reduced implementation errors and improved system robustness. Overall, the system met its functional objectives and demonstrated readiness for practical use as an assistive

CHAPTER 7: CONCLUSION

7.1 Summary of Work Done

The Multimodal Medical Report Analyzer was developed to address the growing complexity and volume of medical reports generated in modern healthcare systems. Medical reports often contain a mixture of unstructured text, structured laboratory tables, and embedded diagnostic images, making manual interpretation time-consuming and dependent on multiple specialist doctors.

This project successfully designed and implemented an end-to-end system capable of processing such multimodal medical reports in an automated and structured manner. The system integrates text extraction with OCR support, laboratory table detection and cleaning, medical image extraction and modality classification, and hybrid rule-based Natural Language Processing for report structuring. A controlled language model was further incorporated to generate patient-friendly and clinician-oriented summaries while maintaining strict safety constraints.

The system was developed incrementally, with individual modules first implemented and validated independently before being integrated into a unified pipeline. Extensive testing on multiple medical reports demonstrated that the system functions reliably across varied report formats and diagnostic domains.

7.2 Key Outcomes and Contributions

The key outcomes achieved through this project include:

• Successful development of a multimodal medical report analysis system capable of handling text, tables, and images within a single framework.

• Implementation of a hybrid rule-based NLP approach that ensures explainability and reliability in structuring medical reports.

• Accurate extraction and classification of embedded medical images based on modality without attempting medical diagnosis.

• Generation of safe and controlled summaries that improve report accessibility for patients and support clinicians in preliminary understanding.

• Reduction of repetitive manual effort involved in reading and organizing medical reports across different medical specializations.

7.3 Limitations of the System

While the proposed system demonstrates effective performance, certain limitations remain. The accuracy of OCR-based text extraction may vary depending on scan quality and document resolution. Additionally, the system focuses on modality identification and report structuring rather than clinical diagnosis or predictive analysis.
The language model component is intentionally constrained to avoid unsafe outputs, which limits the depth of interpretation provided. These limitations were accepted deliberately to prioritize safety, explainability, and ethical responsibility.

7.4 Conclusion

In conclusion, the Multimodal Medical Report Analyzer provides a practical and responsible solution for assisting in medical report interpretation. By combining modular system design, hybrid NLP techniques, computer vision, and controlled language models, the project demonstrates how advanced computing methods can be applied effectively in healthcare without compromising safety or reliability.
The system serves as a supportive tool for healthcare professionals and patients, helping reduce manual workload and improve accessibility of medical information. The successful implementation and validation of the system indicate its potential for further enhancement and real-world adoption.

CHAPTER 8: FUTURE SCOPE

8.1 Enhancements to Image Analysis and Classification

In the current system, the medical image classification module focuses on identifying the modality of extracted images, such as X-ray, CT, MRI, ultrasound, and ECG. In future work, this layer can be enhanced to provide more detailed analysis of classified images.

Advanced deep learning models can be incorporated to extract visual findings from medical images, such as detecting common patterns or anomalies under expert supervision. These enhancements would allow the system to provide richer contextual information alongside image modality classification, while maintaining clear boundaries to avoid unsupervised diagnosis.

8.2 Advanced Natural Language Processing Techniques

The present system employs a rule-based and heuristic NLP approach to structure medical text due to its reliability and explainability. In future versions, this component can be augmented or replaced with machine learning–based NLP models trained on annotated medical text.

Such models could improve adaptability to diverse report formats and complex sentence structures. A hybrid approach combining rule-based logic with machine learning techniques could further enhance accuracy while preserving interpretability and safety.

8.3 Improved Language Model Integration

The explanation and summarization layer currently uses a controlled language model to generate patient-friendly and clinician-oriented summaries. Future enhancements may include integrating more advanced local language models using platforms such as improved Ollama-based deployments.

More capable models can generate clearer, more context-aware summaries while still enforcing strict safety constraints. Fine-tuning or prompt optimization techniques may further improve summary quality without introducing unsafe medical advice.

8.4 Expansion to Additional Report Types

Future development can extend the system to support a wider range of medical report types, including pathology reports, discharge summaries, and longitudinal patient records. This would allow comprehensive analysis across different stages of patient care.

Integration with standardized medical terminologies and coding systems could further improve interoperability with hospital information systems.

8.5 Performance Optimization and Scalability

As the volume of medical data grows, performance optimization will become increasingly important. Future work can focus on improving processing efficiency through batch processing, parallel execution, and optimized resource utilization.

Deployment in cloud or hybrid environments could enable scalable processing of large numbers of reports while maintaining data privacy and security.

8.6 Enhanced User Experience and Accessibility

Future versions of the system may include enhanced visualization techniques, interactive dashboards, and multilingual support to improve accessibility for diverse user groups. Mobile-friendly interfaces could further increase usability for healthcare professionals and patients.

8.7 Ethical and Regulatory Compliance

Future work may incorporate additional safeguards to align with evolving medical regulations and ethical guidelines. This includes stronger audit logs, explainability mechanisms, and compliance with healthcare data protection standards.

CHAPTER 9: REFERENCES

1. Jain, A., & Gupta, D., "A Review on Medical Document Analysis Using OCR and NLP Techniques," *International Journal of Computer Applications*, vol. 182, no. 12, pp. 15–21, 2020.
2. Cohen, P. T., "Natural Language Processing in Healthcare: A Review," *Journal of Biomedical Informatics*, Elsevier, 2019.
3. Smith, J., et al., "Automated Extraction of Laboratory Test Results from Clinical Documents," *IEEE Access*, vol. 8, pp. 145678–145689, 2020.
4. Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.
5. Litjens, G., et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, Elsevier, vol. 42, pp. 60–88, 2017.
6. Tesseract OCR Engine Documentation, https://github.com/tesseract-ocr/tesseract
7. PyMuPDF Documentation, https://pymupdf.readthedocs.io/
8. Camelot: PDF Table Extraction Library Documentation, https://camelot-py.readthedocs.io/
9. Streamlit Documentation, https://docs.streamlit.io/
10. PyTorch Documentation, https://pytorch.org/docs/
11. Pandas Library Documentation, https://pandas.pydata.org/docs/
12. NumPy Library Documentation, https://numpy.org/doc/
13. OpenAI, "Large Language Models and Their Applications," *OpenAI Technical Blog*, 2023.
14. Ollama Documentation, https://ollama.ai/

Kaggle Medical Imaging Datasets

16. Kaggle, *Bone Fracture Detection Dataset*, Kaggle Platform, https://www.kaggle.com/
17. Kaggle, *Brain Tumor MRI Dataset*, Kaggle Platform, https://www.kaggle.com/
18. Kaggle, *Breast Ultrasound Images Dataset*, Kaggle Platform, https://www.kaggle.com/
19. Kaggle, *ECG Images Dataset*, Kaggle Platform, https://www.kaggle.com/
20. Kaggle, *CT Kidney Dataset (Normal, Cyst, Tumor)*, Kaggle Platform, https://www.kaggle.com/
21. Kaggle, *Head CT Hemorrhage Dataset*, Kaggle Platform, https://www.kaggle.com/
22. Kaggle, *Chest X-ray Pneumonia Dataset*, Kaggle Platform, https://www.kaggle.com/
23. Kaggle, *SARS-CoV-2 CT Scan Dataset*, Kaggle Platform, https://www.kaggle.com/
24. Kaggle, *Ultrasound Fetus Dataset*, Kaggle Platform, https://www.kaggle.com/
25. Kaggle, *COVID-19 Radiography Database*, Kaggle Platform, https://www.kaggle.com/
26. Kaggle, *ECG Analysis Dataset*, Kaggle Platform, https://www.kaggle.com/

CHAPTER 10: APPENDIX

A. System Workflow Diagram

This appendix includes the overall workflow diagram of the Multimodal Medical Report Analyzer. The diagram illustrates the complete processing pipeline, starting from medical PDF upload to final structured output and summary generation. It highlights the parallel processing of text, tables, and medical images, followed by hybrid NLP parsing and controlled explanation generation.

(Refer to Figure 4.1 in Chapter 4 for the detailed system workflow.)

B. Sample System Outputs

This section presents representative outputs generated by the system during execution. These samples demonstrate the practical functioning of different system layers and validate the implementation discussed in previous chapters.
The outputs include:

- Extracted textual content from medical PDF reports
- Cleaned laboratory tables with abnormal value indicators
- Structured report sections generated using hybrid NLP parsing
- Classified medical images based on modality
- Patient-friendly summaries generated with safety constraints

These outputs confirm the system's ability to process multimodal medical reports in a structured and reliable manner.

C. Datasets Used

The development and testing of the medical image classification module utilized publicly available medical imaging datasets. These datasets were sourced from the Kaggle platform and cover a wide range of diagnostic modalities, including X-ray, CT, MRI, ultrasound, and ECG images.

The datasets were used strictly for research and educational purposes to validate the system's image classification capability. Detailed dataset references are provided in **Chapter 9: References**.

D. Development Environment

The system was developed and tested in a controlled development environment. Key aspects include:

- Development using Python-based modules
- Initial experimentation and validation in interactive Jupyter Notebook (.ipynb) files
- Final integration and deployment using structured Python (.py) files

- Local execution with a web-based user interface for demonstration and testing

  This development workflow ensured correctness, modularity, and ease of debugging during system implementation.

E. Ethical Disclaimer

The Multimodal Medical Report Analyzer is intended solely as an assistive and informational tool. The system does not provide medical diagnoses, treatment recommendations, or clinical decisions. All generated summaries and outputs are accompanied by disclaimers emphasizing that final interpretation and decision-making must be performed by qualified healthcare professionals.