

Synopsis  
of  
Minor Project [CC3270]

## **AI-Based Health Report Summarizer – MediWay**

**BACHELOR OF TECHNOLOGY**  
in  
Computer Science & Engineering ( Specialisation in IoT and Is )  
Jan – May 2025

By:  
Satyam Nautiyal  
229302154



**MANIPAL UNIVERSITY  
JAIPUR**

Under the guidance of:  
Dr. GL Saini

**Department of Internet of Things and Intelligent Systems**  
**Manipal University Jaipur**

---

## 1. Introduction

In the current healthcare landscape, lab reports ( and diagnostic test results ) are often filled with medical jargon, making it challenging for non-medical professionals to interpret the results. **MediWay** aims to bridge this gap by developing a system that automatically summarizes medical test results into easy-to-understand language narratives, providing users with actionable insights about their health status and save them the anxiety of cluelessness.

---

## 2. Objective

The primary objective of this project is to create an AI-based health report summarizer that:

- Extracts key data from lab reports (in PDF/image format).
- Analyzes the extracted data based on normal reference ranges.
- Provides clear, actionable insights that users can understand without professional assistance.

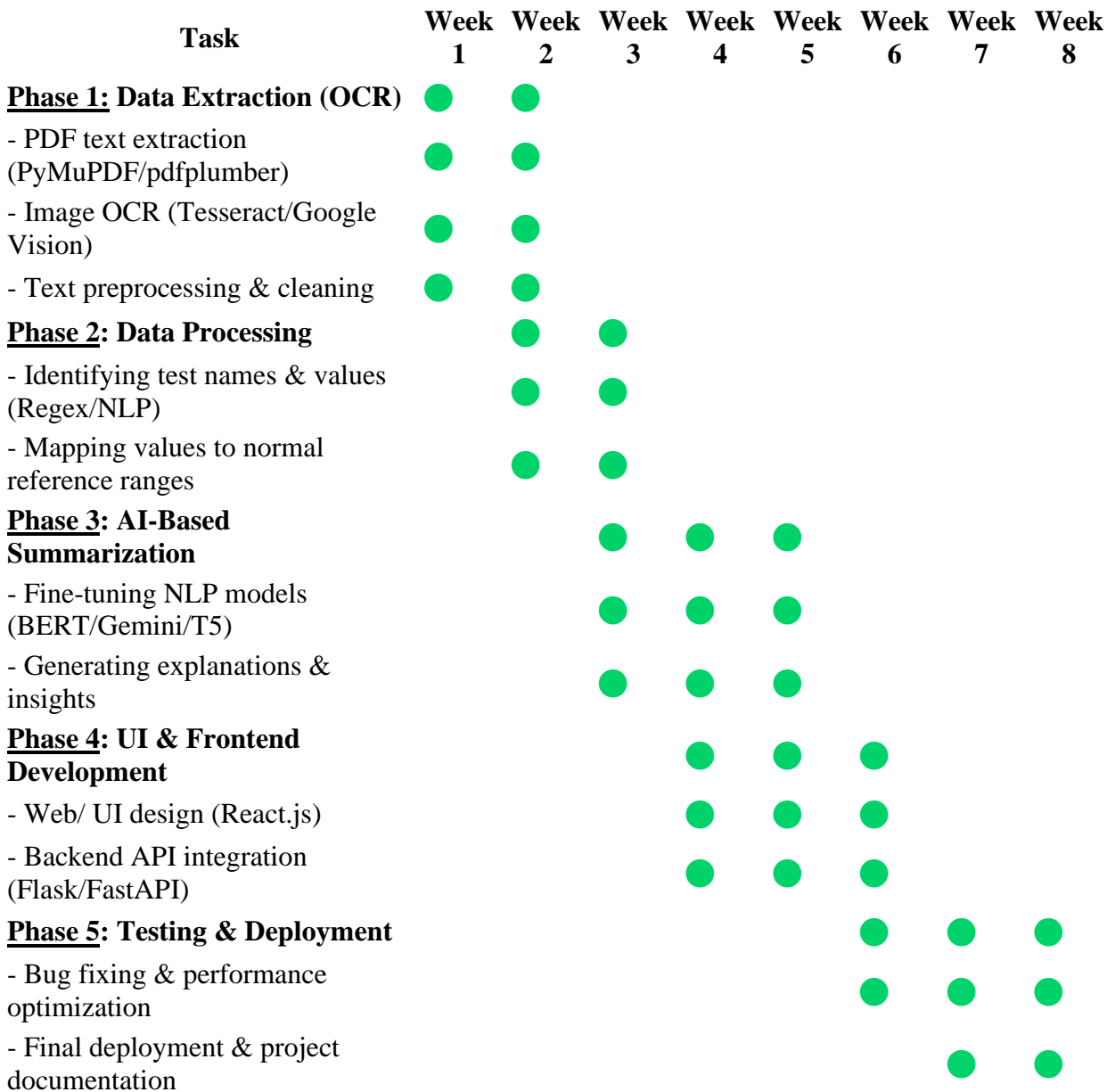
By doing so, MediWay will empower users to interpret their health reports independently and make informed decisions about their health.

---

## 3. Problem Statement

Lab reports and diagnostic test results are often full of medical terms and reference values that may not be easily understood by the average person. Without professional guidance, it is difficult to understand whether certain test results are within the normal range and what actions need to be taken. The challenge lies in transforming technical medical data into a user-friendly format.

#### 4. Gantt Chart for MediWay



● Indicates active work on that task during the given weeks.

---

## 5. Scope

- **Input:** Users will upload their medical reports in PDF or image format.
  - **Processing:** The system will use Optical Character Recognition (OCR) to extract relevant information from the uploaded documents. It will identify test names, results, and compare them with standard reference ranges.
  - **Output:** The system will generate a summary of the results with user-friendly explanations. The summary will indicate whether the test results are within normal ranges and, if not, provide guidance on potential health issues.
- 

## 5. Methodology

- **Phase 1: Data Extraction**
  - Use OCR tools (Tesseract or Google Vision API) to extract data from PDF and image files.
  - Preprocess and structure the extracted data for analysis.
- **Phase 2: Data Processing**
  - Use NLP models (spaCy, BERT) to identify key medical terms and values.
  - Compare test results against predefined normal ranges stored in a database.
- **Phase 3: Summarization**
  - Use pre-trained AI models like **T5** or **GPT** to generate easy-to-understand explanations of the test results.
- **Phase 4: User Interface**
  - Develop a web-based interface to allow users to upload reports and view the results.
- **Phase 5: Future Integration**
  - Once the summarizer is fully functional, integrate a **Doctor Recommendation System** to provide suggestions for specialists based on symptoms.

---

## 6. Technologies to be Used

- **OCR Tools:** Tesseract OCR / Google Vision API
  - **NLP Models:** spaCy, BERT, T5, GPT (Hugging Face Transformers)
  - **Frontend Development:** React.js
  - **Backend Development:** Flask
  - **Database:** SQL / SQLite for storing reference ranges
- 

## 7. Expected Outcomes

- A user-friendly, AI-powered health report summarizer.
  - Easy-to-understand explanations of lab results that help users make informed health decisions.
- 

## 9. Conclusion

**MediWay** will serve as a crucial tool for making complex medical information more accessible to the general public. By leveraging AI and NLP technologies, the project aims to enhance patient understanding of their health, empowering them to take timely action. This project will also pave the way for future healthcare innovations, such as the MediWay Chatbot and Doctor Recommendation System, that integrate seamlessly with health data analysis.

---

## 10. Bibliography / References

### 1. Optical Character Recognition (OCR)

- Smith, R. (2007). "An Overview of the Tesseract OCR Engine." *Proceedings of the Ninth International Conference on Document Analysis and Recognition*. IEEE.
- Google Cloud. "Vision AI – Optical Character Recognition." <https://cloud.google.com/vision>

### 2. Natural Language Processing (NLP)

- Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing*. 3rd Edition. Pearson.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805.

### 3. Machine Learning & AI for Medical Applications

- Rajpurkar, P., et al. (2017). "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." *arXiv preprint arXiv:1711.05225*.
- Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.

### 4. Medical Reference Ranges & Health Data

- Mayo Clinic. (2023). "Laboratory Reference Ranges." <https://www.mayoclinic.org>
- WHO (World Health Organization). "Normal Hematological Ranges and Guidelines." <https://www.who.int>

### 5. Software Development & APIs

- Grinberg, M. (2018). *Flask Web Development: Developing Web Applications with Python*. O'Reilly Media.
- Mozilla. "MDN Web Docs – Web APIs and JavaScript Frameworks." <https://developer.mozilla.org>