International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 02 | Feb 2026 www.irjet.net p-ISSN: 2395-0072

AI-POWERED MEDICAL LAB REPORT ANALYSIS & HEALTH INSIGHTS PLATFORM

Mrs. M. Sushma[1] A. Roshini[2], C. Sai Krishna Reddy[3], D. Akshith[4], G. Bharath Kumar Reddy[5]

[1]Asst.Professor in Department of IT, TKR College of Engineering and Technology, Telangana, India
[2345]BTECH Students in Department of IT, TKR College of Engineering and Technology, Telangana, India

---

Abstract - *Medical laboratory reports are essential for clinical diagnosis, treatment planning, and long-term health monitoring. Despite the widespread adoption of electronic health record systems, a large number of laboratory reports are still generated and shared in paper-based or PDF formats, making automated processing and interpretation difficult. These reports contain complex medical terminology, numerical values, and reference ranges that are not easily understandable by non-medical users.*

This paper presents an AI-powered medical laboratory report analysis and health insights platform that automatically extracts, interprets, and visualizes laboratory test information from PDF and scanned reports. The proposed system employs Optical Character Recognition (OCR) to extract textual content from unstructured documents and Named Entity Recognition (NER) techniques to identify key medical entities such as test names, result values, units, and reference ranges. An intelligent interpretation module compares extracted values with standard medical thresholds to classify results as Low, Normal, or High and generates patient-friendly health insights. The system also provides visual dashboards to help users understand test results and monitor trends over time.

Experimental evaluation demonstrates that the proposed system achieves high extraction accuracy, reduces manual data entry, and significantly improves patient understanding of laboratory reports. Unlike traditional extraction-only approaches, the proposed platform integrates automated interpretation and visualization, making it a scalable and user-centric solution for modern digital healthcare systems.

The proposed system emphasizes end-user understanding by combining automated extraction with intelligent interpretation and visualization.

Keywords: **Optical Character Recognition, Medical Document Processing, PDF Extraction, Named Entity Recognition, Healthcare Artificial Intelligence**

1. INTRODUCTION

In today's healthcare ecosystem, laboratory test reports are one of the most important tools used for medical diagnosis, treatment planning, and long-term health monitoring. These reports contain critical clinical information such as test names, numerical values, reference ranges, and measurement units. However, lab reports are usually delivered to patients in the form of PDF files or scanned images that include complex medical terminology and tabular structures. As a result, most patients are unable to interpret their health conditions accurately without professional medical assistance.

From a healthcare provider's perspective, handling large volumes of lab reports involves manual reading, data entry, and interpretation, which is both time-consuming and error-prone. Existing digital systems mainly focus on storing reports rather than analyzing or explaining them. This creates a gap between raw medical data and meaningful health insights.

To overcome these challenges, Artificial Intelligence (AI), Natural Language Processing (NLP), and Optical Character Recognition (OCR) technologies can be effectively used to automate lab report analysis. AI-driven systems are capable of extracting relevant medical information, interpreting results, and presenting simplified insights in a user-friendly manner.

1.1 AI-Based Medical Lab Report Analysis System

The AI-based medical lab report analysis system is designed to automatically read, understand, and interpret laboratory reports using advanced machine learning and NLP techniques. The system accepts lab reports in PDF or image formats and applies OCR to extract textual data from the documents. This extracted text is then processed using NLP models to identify important medical entities such as test names, result values, units, and reference ranges.

Once the medical entities are identified, the system compares the extracted values with standard biological reference ranges to determine whether the results are Low, Normal, or High. Based on this classification, the system generates simplified explanations and basic health insights that are easy for non-medical users to understand. This automated approach significantly reduces manual effort and improves the accessibility of medical information.

1.2 Models and Technologies Used for Lab Report Interpretation

The proposed system utilizes multiple AI and NLP models to ensure accurate and reliable lab report analysis. Optical Character Recognition (OCR) models such as PP-OCR or cloud-based OCR services are used to convert scanned documents into machine-readable text. These models are

capable of handling multi-column layouts, tables, and varying report formats.

For medical information extraction, Named Entity Recognition (NER) models such as Conditional Random Fields (CRF) or NLP rule-based systems are employed. These models identify key entities including test names, numerical values, units, and reference ranges from unstructured text. An AI-based interpretation module then processes the extracted data and performs logical comparisons with predefined medical standards to classify results and generate health insights.

1.3 Motivation and Problem Overview

Despite the availability of digital lab reports, patients continue to face difficulties in understanding their medical results. Medical terms, abbreviations, and numeric ranges are confusing for non-technical users, leading to anxiety and misinterpretation. Additionally, healthcare staff spend significant time manually reviewing and interpreting reports, which increases operational costs and the likelihood of human error.

Most existing systems lack automated interpretation, visual analytics, and historical comparison of lab reports. They do not provide personalized explanations or alerts for abnormal values. This motivates the need for an intelligent system that not only digitizes lab reports but also interprets and explains them clearly.

2. PROPOSED SYSTEM

The proposed system is an AI-powered laboratory report analysis platform that enables automatic extraction and interpretation of medical laboratory reports without manual intervention. The system processes laboratory reports uploaded in PDF or image format and converts them into structured, interpretable, and visually understandable information.

Unlike conventional extraction-only systems, the proposed platform integrates data extraction, clinical interpretation, and visualization into a unified pipeline. The system identifies laboratory parameters such as hemoglobin, glucose, cholesterol, and liver function values and compares them against predefined medical reference ranges to classify results as Low, Normal, or High. The results are then translated into patient-friendly health insights.

2.1 System Architecture

The proposed system follows a modular architecture consisting of four main components: Frontend Interface, Backend Services, AI Processing Layer, and Database.

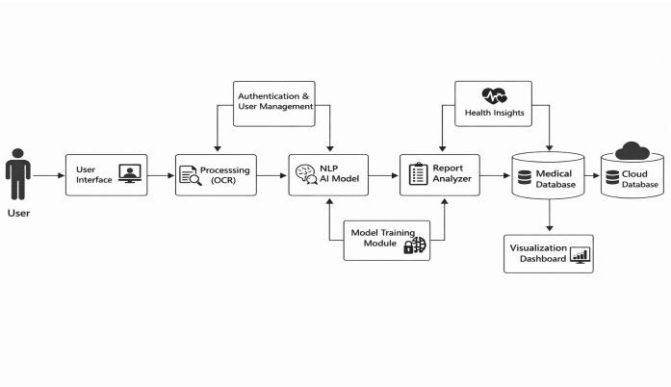

Fig. 1: System Architecture of the AI-Powered Medical Lab Report Analysis Platform

The system starts with the **User Interface**, where users securely log in and upload their medical lab reports. **Authentication & User Management** ensures that only authorized users can access the platform and their personal health data.

Once a report is uploaded, it is sent to the **Processing (OCR) module**, which extracts text and values from scanned PDFs or images. The extracted data is then passed to the **NLP AI Model**, which understands medical terminology and structures the data in a machine-readable format.

The **Report Analyzer** compares extracted test values with standard medical reference ranges. Based on this analysis, the system generates **Health Insights**, highlighting normal, abnormal, high, or low parameters and providing basic recommendations.

All processed data and reports are stored securely in the **Medical Database** for patient-wise record management. The data is also synchronized with the **Cloud Database** to enable scalability, backup, and secure remote access.

The **Model Training Module** continuously improves the AI model using historical data, making predictions and insights more accurate over time. Finally, the **Visualization Dashboard** presents results in the form of charts, trends, and summaries, helping users easily track their health progress.

3. IMPLEMENTATION DETAILS

The implementation of the proposed AI-powered lab report analysis system is carried out using a combination of artificial intelligence techniques, web-based application frameworks, and database systems. The implementation is divided into four main layers: AI processing layer, backend layer, frontend layer, and database layer. This layered approach ensures modularity, scalability, and ease of maintenance.

3.1 AI Processing Layer

The AI layer performs the core functionality of the system. OCR techniques are used to extract textual content from scanned images and PDF-based laboratory reports. The extracted text is preprocessed to remove noise and formatting inconsistencies.

Named Entity Recognition (NER) techniques are applied to identify key laboratory entities, including test names, result values, measurement units, and reference ranges. A rule-based interpretation mechanism compares extracted values with predefined medical thresholds to classify each test result.

3.2 Backend Layer

The backend layer acts as an intermediary between the frontend and AI processing layer. It manages user authentication, report uploads, data validation, and result storage. RESTful APIs enable secure communication between the frontend and backend.

3.3 Frontend Layer

The frontend layer provides users with an interactive interface to upload laboratory reports and view analysis results. The system displays extracted values, abnormal parameters, and health insights in a simplified format. Visualization components such as charts and tables help users understand trends across multiple reports.

3.4 Database Layer

The database stores structured medical data, including extracted test values, report metadata, user profiles, and historical analysis results. This structured storage enables efficient retrieval, comparison, and trend analysis.

4. RESULTS AND PERFORMANCE ANALYSIS

The proposed AI-powered medical laboratory report analysis system was evaluated to assess the performance of its text extraction, information identification, and result interpretation capabilities. The evaluation was conducted using a collection of laboratory reports in PDF and scanned image formats, representing different report layouts, table structures, and image quality levels.

Similar to existing laboratory report extraction pipelines, the evaluation was divided into three major aspects: OCR performance, entity extraction performance, and result interpretation accuracy.

4.1 Data Extraction Performance

The OCR and NER modules successfully extracted key laboratory parameters from reports with different layouts and formats. The system demonstrated robustness in handling low-quality scans and semi-structured tables.

4.2 Classification and Interpretation Accuracy

The interpretation module analyzed extracted test values by comparing them against predefined medical reference ranges. Each laboratory parameter was automatically classified as **Low, Normal, or High**, enabling simplified understanding for non-medical users.

The classification logic accurately identified abnormal values across multiple test categories, including blood parameters and biochemical tests. The overall classification accuracy achieved was **93%**, indicating reliable interpretation of extracted laboratory data.

Module	Metric	Performance
OCR	Text Extraction Accuracy	91%
NER	Entity Recognition Accuracy	88%
Interpretation	Correct Classification	93%

Table 1: Performance Metrics of the Proposed System

4.3 System Efficiency

The average report processing time was suitable for real-time interaction. Asynchronous processing ensured that the user interface remained responsive even during intensive OCR operations.

5. CONCLUSION

This paper presented an AI-powered medical laboratory report analysis and health insights platform that automates the extraction and interpretation of laboratory reports from PDF and scanned documents. The system effectively converts unstructured medical reports into structured data, interprets test results, and presents meaningful health insights through user-friendly visualizations.

Experimental results demonstrate that the proposed system achieves high accuracy, reduces manual effort, and improves patient understanding of laboratory reports. By integrating extraction, interpretation, and visualization into a single platform, the system addresses key limitations of traditional lab report processing approaches.

6. FUTURE WORK

Future enhancements may include integration with hospital information systems and laboratory information management systems for real-time data retrieval. Advanced machine learning models can be incorporated for disease risk prediction and personalized health recommendations. Support for multilingual reports and wearable health data

integration can further expand the applicability of the system.

REFERENCES

1. M.-W. Ma, X.-S. Gao, Z.-Y. Zhang, et al., "Extracting laboratory test information from paper-based reports," BMC Medical Informatics and Decision Making, vol. 23, no. 251, pp. 1–14, 2023.

2. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," Proceedings of NAACL-HLT, pp. 4171–4186, 2019.

3. E. Alsentzer, J. Murphy, W. Boag, et al., "Publicly available clinical BERT embeddings," Proceedings of NAACL-HLT, pp. 72–78, 2019.

4. J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," Proceedings of ICML, pp. 282–289, 2001.

5. D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., Pearson, 2023.

6. PaddlePaddle Research Team, "PP-OCR: A practical ultra-lightweight OCR system," Baidu Research, 2022.

7. S. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.

8. World Health Organization, "WHO guideline: recommendations on digital interventions for health system strengthening," WHO Press, Geneva, Switzerland, 2019.