WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

Smart Medical Report Analyzer: A Software Solution for Easy Reading of Laboratory Reports

Prof. D. A. Kapgate[1], Sujal Dupare[2], Prerna Kumley[3], Shivpriya Deshmukh[4]

Department of Electronics and Communication Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India

ABSTRACT-

Medical reports include test names, numerical measurements, units, and reference ranges that may be challenging to interpret for a non-expert. This article describes the design of a Medical Report Analyzer which can sort uploaded laboratory reports and generate an easy to interpret summary. The suggested workflow accepts a PDF file or image, extracts the text using OCR when needed, identifies test names and corresponding values, and compares each value to the interval printed on the same report. Possible low, in-range or high flags and brief explanations are presented. The tool is solely for information support, it should not be used to diagnose disease or determine treatment. Given that report layouts and reference intervals differ, extracted values must be attributable to the original report, and indeterminate values must be checked manually. Accurate statements should only be made after validation of tagged sample reports.

Index Terms- Medical Report Analyzer, laboratory reports, optical character recognition, extraction of information, reference range, summarizing a report.

I. INTRODUCTION

Laboratory tests yield data that aids clinicians in monitoring a patient's health. A report can have various fields, such as test names, numeric values and units, reference ranges and label abbreviations. A natural language processing utility that sorts these fields and can translate them into natural language could assist a user in reviewing a report.

Never use a laboratory result in isolation. Reference ranges can vary between laboratories and even testing methods, and a value outside the printed range may not mean the disease is present, while a value within the range may not mean the animal is healthy [1]. That is, the analyzer should (1) state what the information mean in a report, (2) make the comparison view clear and (3) it is hard to make the automated output a diagnosis. Previous report digitization researches are similar to the general workflow combining optical character recognition (OCR) and information extraction processes to extract test name, result, units and reference values [2]. The project applies it to the user presentation and clarify the source value and the uncertain extraction.

II. PROPOSED SYSTEM

A. System Design

The system will include a report upload interface, document-reader, result- extraction-and-normalisation module, reference-range-compared-module, and summary-view. Selectable-text PDF can be parsed directly, otherwise OCR is needed for pictures and graphics. Extracted entries are stored in a standard format, independent of document originating, providing the testname, observed-value, unit, printed-reference-interval, and extraction-success. The analyzer compares compatible numerical values to the interval printed on the report, and produces a human-ready summary.

Upload report (PDF / Images)

↓

Text Extraction or OCR

↓

Test name, value, units, and range parsing

↓

Validation and Reference-Range Comparison

↓

Flagged Results + Plain-Language Summary

B. Main Functions

Sample input: the user chooses their laboratory report from a device Visualisation: the interface should show what types of files it can accept and should give user feedback if something cannot be parsed Data extraction: the system determines which report fields are meaningful and associates each measured value with the correct test, unit and interval Normalisation: spacing, common symbols, number formatting, etc. should be made consistent and applied without automatically altering the value reported Result flagging: if there is a valid numeric range the value should be tagged as below, within or above that particular range Summary generation: the results are presented to the user in table form, with simple descriptions and a reminder to see a clinician for interpretation

C. Safety, Privacy and Traceability

Just try to keep the original report on the screen, so that the user can check that each value has been extracted. Use low-confidence OCR output, missing units, ambiguous rows and unrecognised intervals should be annotated as uncertain (and not guessed). Remove personal identifiers, retain the uploaded file for only as long as necessary, and impose access controls on stored reports. Do not recommend medicines, treatment changes or diagnose based on individual values.

III. METHODOLOGY

A. Upload Report and Extract Text

This process starts when the user uploads a PDF or an image. PDF text can be extracted directly, using OCR for scans. For scanned pages, recognition can be affected by quality, skew, contrast, handwriting, narrow columns, and small fonts. All extracted entries should carry a reference to the page to provide end users with an opportunity for comparison.

B. Field Identification and Normalization

Text is extracted into candidate row/field sorts and selected by the parser test name, result, units, and reference intervals, including high and low symbols from the laboratory. Fields are grouped into one record. Normalization should keep the decimal points, inequality signs, leading negative signs, and unit prefixes. If a value or relationship of a field to other values is not immediately obvious, return that entry for manual processing.

C. Reference-Range Comparison

Comparison is only performed if the measured value and range can be meaningfully compared and are in comparable units of measurement. Only the interval printed on the uploaded report is considered; it is given precedence over the range stored in the system as laboratories use different methods and different intervals [1]. For a numeric range with lower limit L, upper U and measurement x, one can define the following numeric flag: $x < L$ (below range), L x U (within range) and $x > U$ (above range). This flag only indicates whether the value is within the displayed interval, it gives no information about the likelihood of disease or the need for urgent action.

D. Summary and User Interface

The following fields are to be displayed on the result screen: test name, value extracted, unit, printed interval, flag, short plain language note. High/low indicators provided by the laboratory can be used with source shown. Summary must clearly state: this is not a diagnosis, can be corrected and recreated, simple interface for people with varying levels of digital literacy.

IV. HARDWARE AND SOFTWARE REQUIREMENTS

Component / Tool	**proposed function**
Computer or smartphone	Upload report and view report
Document reader / OCR	Read PDFs or extract text from images
Field extraction	Test name, value, unit, and range to be determined.
Validation rules	Keep units and mark fields as uncertain
Range comparator	Compare to the period on the report
Summary interface	Show flags and plain-language notes
Privacy controls	Reduce the data stored and to allow access

V. VALIDATION AND EVALUATION

The system should be evaluated on de-identified sample reports that utilize a variety of layouts, font types, image qualities and interval formats. A manually checked transcription will act as the 'truth'. The accuracy of text extraction can be measured using error rates at the character or word level. Extraction of test names, values, units and intervals can also be measured in this way, as well as separately for each flag, value and interval.

Usability testing should verify users can find a result, interpret its flag, and tell an auto-explanation from clinical guidance. No accuracy, response-time, or usability scores are claimed herein because project-specific metrics are not offered. Future performance claims should include the dataset, method, and results.

VI. LIMITATIONS AND FUTURE WORK

Scans that are too blurry, excessive use of hand writing, table and layout issues, and poor handwriting will cause the system to make errors. Not all results are numerical. Reference ranges are not always the same for everyone. In the future, the system may be further extended to report confidence scores, automatic annotation, correction, a verifiable history, in a number of languages, and additional testing. Any clinical application would require validation, privacy controls, and manual review.

VII. CONCLUSION

This document explains a plan for Medical Report Analyzer that offers Laboratory value organization and plain language explanation. Report upload, text extraction, field parsing, conservative normalization, comparison with printed reference intervals. Traceability, privacy, and clear boundaries are critical: output helps reading and information organization, not diagnosis and therapy. Reliability need to be identified with the validation of reports marked by humans.

ACKNOWLEDGMENT

Authors are thankful to their project guide Prof. D. A. Kapgate and the members of Department of Electronics & Communication Engineering, Priyadarshini College of Engineering, Nagpur.

REFERENCES

[1] You, J., & Shin, H. (2026). Large Language Models for Interpretation of Health Checkup Results. npj Digital Medicine, 9, 565.

[2] Bhasuran, B., et al. (2025). Preliminary Analysis of the Impact of Lab Results on Large Language Model Generated Differential Diagnoses. npj Digital Medicine, 8, 166.

[3] Cadamuro, J., et al. (2023). Potentials and Pitfalls of ChatGPT and Natural-Language Artificial Intelligence Models for the Understanding of Laboratory Medicine Test Results. Clinical Chemistry and Laboratory Medicine.

[4] Bhasuran, B., et al. (2026). Evaluation of Causal Reasoning for Large Language Models in Contextualized Clinical Scenarios of Laboratory Test Interpretation. npj Digital Medicine, 9, 487.