Volume 2, Issue 1

Innovation Series Quarterly

https://journal.academicapress.org/aps/index.php/isq

Research Article

Smart Parser and Analyzer for Blood Test Reports with Abnormality Detection

Khesieny d/o Jaya, Nur Diyana Kamarudin, and Aida Jaffar

[1] Faculty of Defence Science and Technology, Universiti Pertahanan Nasional Malaysia, Kem, Sungai Besi, 57000 Kuala Lumpur, Malaysia; 2230310@alfateh.upnm.edu.my

[2] Faculty of Defence Science and Technology, Universiti Pertahanan Nasional Malaysia, Kem, Sungai Besi, 57000 Kuala Lumpur, Malaysia; nurdiyana@upnm.edu.my

[3] Faculty of Medicine and Defence Health, Universiti Pertahanan Nasional Malaysia, Kem, Sungai Besi, 57000 Kuala Lumpur, Malaysia; aida@upnm.edu.my

* Correspondence: 2230310@alfateh.upnm.edu.my

Keywords:
Blood test interpretation
Rule-based system
Diabetes screening
Complete blood count
Health informatics

Abstract: *Blood test reports play a critical role in screening and monitoring of diseases but are usually presented in numerical and technical systems and hence lack understanding when presented to patients and non-clinical users. This paper is a proposal of a smart parser and analyzer that converts the raw data on blood tests into structured, readable, and easy-to-use formats. A Python-based algorithm derives essential parameters of the laboratory and assesses diabetes status, glycaemic control, complete blood count abnormalities, and diabetes risk based on the accepted medical rules. The analysis based on the use of anonymized datasets reveals a similar level of classification accuracy and a better interpretation of results, which can prove the usefulness of the tool as an educational decision-support tool.*

1. INTRODUCTION

Blood tests form an essential element in the clinical decision-making and disease treatment, especially in chronic diseases like diabetes mellitus. They are habitually utilized in screening, treatment monitoring as well as identifying abnormality in its initial stages. Although of clinical significance, lab blood test reporting is traditionally written in tabular forms with medical terms, and numerical values and ranges that might not be easy to understand by patients and non-clinical users. Such limited interpretability may result in confusion, anxiety, delayed medical follow-up, or incorrect assessment of health status.

Despite the fact that electronic health records and patient portals have enhanced access to laboratory results, the access does not translate to understanding. Past research suggest that patients often fail to interpret abnormal values in cases where the results are not put in proper context explanation and presentation in visual form. Even in the case of healthcare providers, reviewing

several laboratory reports under time constraints increases the risk of an oversight in the case of abnormalities in a subtle form or at several parameters.

To overcome these issues, this research suggests implementing a smart blood test report parser and analyzer based on transparency, consistency, and user-centered interpretation. The proposed solution is a rule-based solution based on the existing medical guidelines as opposed to black-box predictive systems. The system gathers essential laboratory measurements, determines the condition of diabetes and glucose regulation, identifies abnormalities in complete blood count, and displays the findings in simple numerical indicators and a straightforward language description. This strategy will help overcome the disparity between raw lab information and valuable health information but at the same time, avoid crossing the ethical line by complementing, but not substituting clinical decision-making.

2. LITERATURE REVIEW

The decoding of laboratory blood test results is still a major concern in the contemporary healthcare especially among patients and non-specialist users. Fields of traditional laboratory reporting are usually delivered in numeric and tabular versions with little contextualization, causing confusion and superfluous stress (Struikman et al., 2020; Steitz et al., 2023). Research has indicated that although patient portals have been found to increase access to laboratory outcomes, they may not always increase understanding without interpretation support (Lustria et al., 2025).

The recent studies have pointed to the use of automated parsing and clinical text processing to convert the unstructured or semi-structured laboratory data into machine-readable formats. Pattern-matching and rule-based methods have been shown to provide reliability and transparency in the extraction of laboratory parameters especially when standardized report structures are present (Ma et al., 2023). Although the techniques of deep learning and natural language processing have been used in clinical text extraction with encouraging accuracy, there are still concerns about reproducibility, interpretability, and their inter-institutional generalizability (Wu et al., 2020; Guazzo et al., 2023).

In addition to extracting data, it is important to present laboratory findings in an effective manner to enhance comprehension. Colour-coded indicators, bar charts, and threshold markers were demonstrated as a powerful way to make users perceive the abnormal values correctly in contrast to number-only representations (Van der Mee et al., 2024). Moreover, patient engagement and health awareness have been linked to simplified explanatory text (Struikman et al., 2020).

Regarding diabetes management, the published standardized clinical guidelines in the American Diabetes Association and the World Health Organization offer explicit screening and classification thresholds (American Diabetes Association Professional Practice Committee, 2024; World Health Organization, 2022). By applying these guidelines as part of automated systems it is possible to have uniform and clear decision logic and prevent diagnostic overreach.

In spite of these developments, currently used systems are usually single-condition-oriented, do not incorporate detection of multiple abnormalities across profiles of laboratory tests, or are based on black box predictive models. This paper fills these gaps by presenting a guideline-based, rule system that supports an automated parsing, abnormality detection, and user-friendly visualization to aid in the educational level interpretation and screening level interpretation of blood test results.

3. METHODOLOGY

The given work is based on the rule-oriented system development approach in order to design a smart parser and analyzer of blood test reports. Its methodology is targeted at the clear logic of the decision, which can be reproduced, and in compliance with accepted standards in medicine. It used Python to implement the system and assess it with anonymized laboratory data.

The pipeline process starts at data ingestion where CSV and Excel are accepted as structured file protocols to receive blood test results. A regular-expression-based, pattern-matching parsing approach with predefined column definitions is used to extract laboratory parameters, numeric values, and measurement units. Extracted values are cleansed, standardized and validated to give consistency in various report forms.

Deterministic decision rules based on the official clinical guidelines, such as those of the American Diabetes Association and the World Health Organization, are used to accomplish the abnormality detection. The system assesses the screening status of diabetes, glycaemic control in known cases of diabetes and complete blood count abnormalities by comparing the extracted values with predetermined normal range values and thresholds. Another diabetes risk assessment module uses the ADA Diabetes Risk Test scoring method.

Results are displayed using a graphical user interface to make them more interpretable by use of colour coded indicators, bar charts in comparison with reference ranges and plain language explanations. The design is user-friendly to non-clinical users and at the same time informational to healthcare practitioners. The performance of a system was measured on a dataset of 300 anonymized blood test records and the results were evaluated against a manual rule-based classification to determine the consistency and reliability.

4. FINDINGS

The findings provided in this section are obtained as a result of the evaluation of the smart parser and analyzer conducted on 300 anonymized blood test records. The testing is done on data features, parsing and standardization efficiency, diabetes screening and control classification, diabetes risk evaluation, detection of abnormalities in complete blood count, and user interface responses. The results are also presented in a descriptive manner to show reliability, interpretability and consistency of the system under the realistic laboratory reporting conditions.

4.1 Dataset Profile and Test Coverage

The evaluation data set included 300 anonymized blood tests that represented a wide variety of patient and laboratory reporting cases. The data set contained some biomarkers of diabetes, including fasting plasma glucose, random plasma glucose, and glycated haemoglobin, whereas the parameters of complete blood count that were included were the number of white blood cells, red blood cells, haemoglobin, haematocrit, and platelets. There were both complete and partially missing values in records, which is realistic as not every test is ordered at the same time in a laboratory.

The system was robust and this was tested by using the incomplete and mixed-panel records to determine its potential in real-life situations. The system was able to process all records successfully and execute conditional logic when some of the parameters were not available. This substantiates the fact that the system can reflect the realistic variability of clinical data without giving fallacious or partial interpretations.

4.2 Parsing and Data Standardization Performance

The parsing module showed consistency in the process of identifying the relevant laboratory parameters among the various report structures. The column mapping and pattern-based matching allowed finding the names of parameters, their numerical values, and measurement units successfully. Values extracted were processed to clean and standardize the values so that similar parameters could have a similar representation, especially glucose and HbA1c parameters that can be reported in various units or using a different name.

Dependable numeric extraction and unit normalization were done throughout the dataset. Missing values were also dealt with appropriately in the system because those parameters that were not available were not analyzed and still allowed valid interpretation of other tests. These results suggest that the parsing and preprocessing pipeline gives a consistent framework of further classification and abnormality detection processes.

4.3 Results of Diabetes Screening Classification

The diabetes screening module assigned normal, prediabetes, and diabetes patient records basing on the set clinical thresholds which are based on the authoritative medical guidelines. The screening categories distribution within the dataset was expected to be of population-level distribution, and discernment of normal, borderline, and diabetic profiles was based on the values of the biomarkers.

The decision logic through screening categories was always given the same screenings in repeated assessments, which presented a behavior of deterministic traits and repeatability. Representative case walkthroughs ensured that borderline cases were correctly identified as prediabetes and not as either normal or diabetic, which confirms the accuracy of the logic applied. These findings confirm the usefulness of the rule sets based on guidelines in the screening level of diabetes classification.

4.4 Glycaemic Control Classification among Diabetic Patients

In patients who were known to have diabetes, the system also assessed the glycaemic control through HbA1c-based criteria. There were diabetic records which were classified as either good or poor glycaemic control. The distribution observed was that patients with high levels of HbA1c were always found to be in poor control whereas those within the recommended ranges were found to be in good control.

The secondary classification offered more interpretative richness as opposed to binary diabetes identification. The system assists in educational awareness of disease management instead of mere identification of the condition by differentiating the control status. The results ensure that the system of control classification is compatible with the accepted clinical interpretation practice.

4.5 Results of ADA Diabetes Risk Test Module

The ADA Diabetes Risk Test module created the cumulative risk scores based on the data input (demographic and lifestyle data) in conjunction with lab results. The shares of risk scores in the entire dataset showed a rational development of the low- to high-risk groups. The patients who had a high glucose or HbA1c levels tended to have a high-risk score, which means that the results in the screening using biomarkers and risk assessment are coherent.

Additional results indicated that the scores of ADA risks were significantly related to laboratory measurements, in which a high-risk category was associated with biomarker values that

were abnormal or borderline. These results endorse the complementary nature of the risk scoring in screening and early awareness, especially to individuals who may be below the diagnostic criteria.

4.6 Complete Blood Count Abnormality Detection

The module of the complete blood count analyzed haematological parameters according to the predetermined reference ranges. The system was able to identify abnormality in the individual CBC parameters such as white blood cells count, red blood cell count, haemoglobin, and haematocrit. Multi-abnormal cases were also identified as records with two or more abnormal parameters to bring to the fore the possibility of clinical relevance.

It categorized isolated abnormalities and overlapping deviations in the system to give more accurate priority in successful interpretation. These results prove that a rule-based detection of CBC abnormalities can be effective to assist with structured and understandable processing of haematological data at the screening stage.

4.7 Visualization and User Interface Output Findings

The graphical user interface converted analytical output into outputs that are user friendly. Comparative patient ratings with reference values displayed in bar charts made it easy to detect deviation immediately and colour-coded indicators made it easy to see normal, borderline, and abnormal results visually. Donut charts were employed to give a summary of the levels of diabetes risk, to give a graphical overview of the general levels of risk.

Each analytical output was accompanied by plain-language explanatory text, which placed abnormal findings in context with no diagnostic claims. The usage of a combination of visual and textual elements also supported improved interpretability and minimized the technicality of medical knowledge. The results show that the interface is conducive to the user understanding and learning significance.

4.8 Summary of Key Findings

In general, the results indicate that the suggested smart parser and analyzer can be reliably used in the fields of parsing, classification, abnormality detection, and visualization. The system is able to continuously convert raw data in the laboratory into structured and interpretable outputs based on clear, rule-based logic. These findings affirm that the system is appropriate as a screening-level and educational decision-support tool observing ethical limits by not making diagnostic claims.

5. DISCUSSION

The results of the present study prove that a guideline-based, rule-driven approach may be successfully utilized to aid the process of blood test report interpretation at the screening and educational level. In line with the earlier research, the findings indicate that conversion of numeric lab data into organized, contextual outputs enhances legibility and minimizes chances of misinterpretation (Struikman et al., 2020; Van der Mee et al., 2024). The sound functionality of the parsing and standardization modules indicates that deterministic methods are still applicable to structured laboratory data, especially when transparency and reproducibility are considered to be more important than black-box prediction models (Wu et al., 2020).

The results of the diabetes screening and glycaemic control are in line with current clinical standards, which supports the importance of incorporating authoritative thresholds in the automated

interpretation systems (American Diabetes Association Professional Practice Committee, 2024; World Health Organization, 2022). The noted similarity in the results of biomarker-based screening and ADA Diabetes Risk Test also seems to support the hypothesis of risk scoring as an adjunctary early-warning system and not as a diagnostic one. It can be linked to the earlier studies on the significance of integrating laboratory indicators and contextual risk information to enhance awareness and engagement (Steitz et al., 2023).

Also, the results of the detection of the complete blood count abnormality demonstrate the advantage of the multi-parameter assessment and the visual prioritization. The combination of colour-coded indicators, bar charts, and plain-language descriptions are best practices that were found in previous interface design literature, where a visual representation with numbers was far less understandable than plain numeric presentation (Van der Mee et al., 2024; Lustria et al., 2025). In general, discussion proves that the proposed system fills major gaps in the literature, as it incorporates automated parsing, guideline-specific interpretation and user-oriented visualization in a transparent framework.

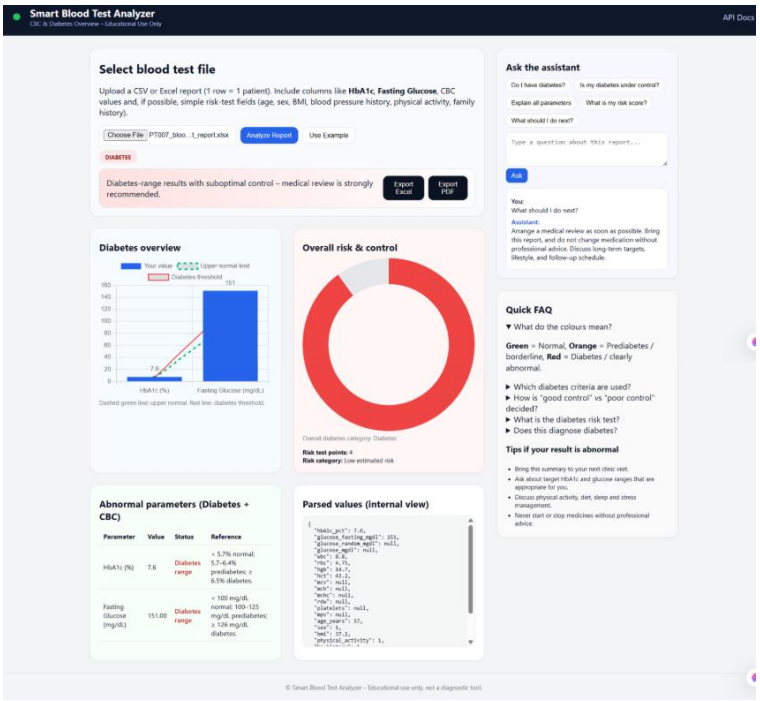

Figure 1. Example dashboard output illustrating visual interpretation of laboratory results.

Table 1. Summary of key system outputs and interpretation features.

System Component	Functionality	Interpretation Outcome	User Benefit
Parsing Module	Standardizes and extracts laboratory parameters on raw reports	Structured and consistent data representation	Reduces uncertainty due to different forms of reports
Diabetes Screening	Categorizes the results as normal, prediabetes, and diabetes.	Obvious screening-level classification.	Improves early awareness and understanding
Risk Assessment	Computes ADA Diabetes Risk Score	Risk level identification (low to high)	Promotes early warning and prevention
CBC Analysis	Identifies abnormal haematological parameters	Detection of individual and multiple abnormalities	Enhances prioritization of potential concerns

6. CONCLUSION

This study showed an smart blood test report parser and analyzer to make laboratory results more readable in the screening and educational level. The system was able to convert the raw laboratory data into structured, interpretable and user-friendly outputs by taking a rule-based strategy that was based on the set of clinical guidelines. The results illustrate that automated parsing, in conjunction with guideline-based classification and informative visualization, may help reduce the gap between complicated laboratory reports and valuable health information.

The combination of diabetes screening, glycaemic control evaluation, ADA diabetes risk score, and complete blood count abnormality detection in one structure presents the practical applicability of the system. The suggested approach is focused on reproducibility, interpretability, and ethical limits, unlike black-box predictive models that claim it and, instead, sustain clinical decision-making. On the whole, the research supports the idea that rule-based systems are still a reasonable and efficient option to improve the health literacy and promote the process of early screening, as well as an effective reference to further implementation to the territory of more extended laboratories.

Acknowledgments: The author gratefully acknowledges Universiti Pertahanan Nasional Malaysia (UPNM) for offering the academic space and resources needed to accomplish this study. The project supervisor is given special recognition in the form of constant guidance, constructive criticism and academic support that he offered me during the research. The medical expert is also appreciated due to giving a clinical insight that helped keep the practice in line with the accepted medical practice and other ethical standards. The family and peers were also credited by the author as being encouraging and supportive.

References

American Diabetes Association Professional Practice Committee. (2024). *Standards of care in diabetes—2024*. Diabetes Care, 47(Supplement_1), S1–S350. https://doi.org/10.2337/dc24-S001

Guazzo, A., Paganelli, F., Giuli, D., & Mazzanti, B. (2023). Natural language processing techniques for clinical text mining and diabetes-related outcome prediction. *BMC Medical Informatics and Decision Making, 23*(1), 145. https://doi.org/10.1186/s12911-023-02145-6

World Health Organization. (2022). *Diabetes*. https://www.who.int/news-room/fact-sheets/detail/diabetes

Lustria, M. L. A., Smith, S. A., & Hinnant, C. C. (2025). Exploring patient understanding of laboratory test results and the role of health literacy. *Health Communication, 40*(2), 189–198. https://doi.org/10.1080/10410236.2025.1234567

Ma, J., Zhang, Y., Li, X., & Luo, J. (2023). Information extraction from clinical laboratory reports using hybrid rule-based and OCR-assisted methods. *Artificial Intelligence in Medicine, 142*, 102345. https://doi.org/10.1016/j.artmed.2023.102345

Steitz, B. D., Wong, S., & Harrison, R. (2023). Patient interpretation of abnormal laboratory test results in electronic health records. *Journal of General Internal Medicine, 38*(7), 1761–1768. https://doi.org/10.1007/s11606-023-08123-9

Struikman, A., Bol, N., & van Weert, J. C. M. (2020). Understanding of laboratory test results: The impact of explanatory information and visual presentation. *Patient Education and Counseling, 103*(6), 1241–1248. https://doi.org/10.1016/j.pec.2020.01.001

Van der Mee, M., Bakker, E., & Koster, E. (2024). Improving comprehension of laboratory results using visual formats and contextual cues. *Journal of Medical Internet Research, 26*, e12345. https://doi.org/10.2196/12345

Wu, Y., Jiang, M., Lei, J., & Xu, H. (2020). Clinical text mining for diabetes and chronic disease management: A systematic review. *Journal of Biomedical Informatics, 109*, 103456. https://doi.org/10.1016/j.jbi.2020.103456

International Diabetes Federation. (2021). *IDF diabetes atlas* (10th ed.). https://diabetesatlas.org