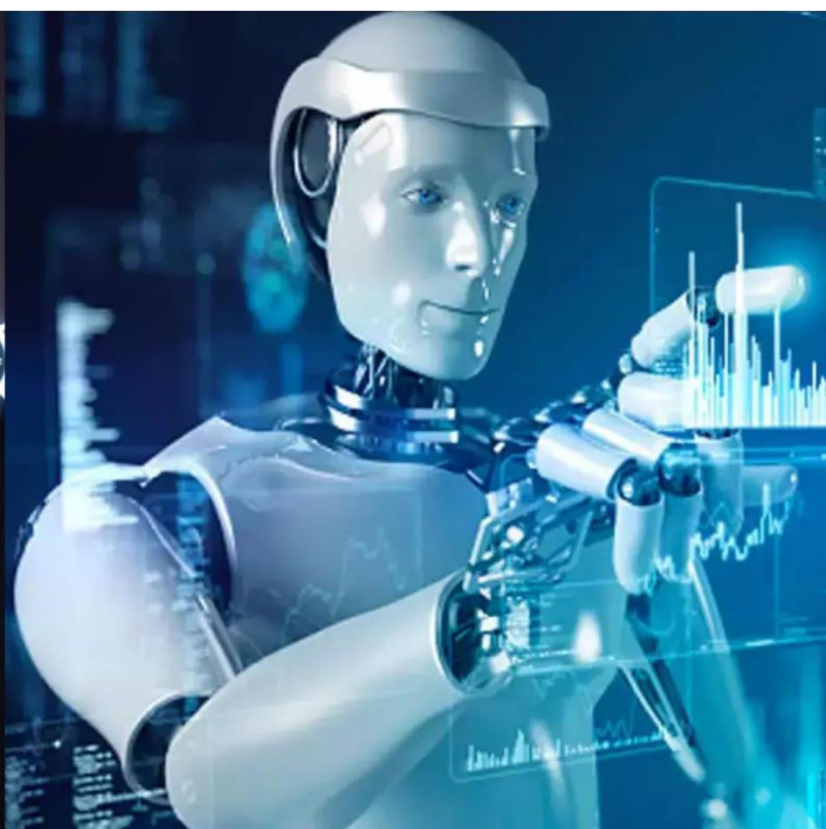




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 10, October 2026

MediSimplify AI: An Intelligent Medical Report Simplification System

A.Gayathri, Avina T. S, V. Karpagam

U.G. Student, Department of Computer Science and Engineering, K.L.N. College of Engineering, Madurai,
Tamil Nadu, India

U.G. Student, Department of Computer Science and Engineering, K.L.N. College of Engineering, Madurai,
Tamil Nadu, India

Associate Professor, Department of Computer Science and Engineering, K.L.N. College of Engineering, Madurai,
Tamil Nadu, India

ABSTRACT: Medical reports contain technical terminology, abbreviations, numerical values and reference ranges that may be difficult for patients without a medical background to understand. This paper presents MediSimplify AI, an intelligent medical report simplification system designed to convert report content into patient-friendly explanations while preserving the original clinical information. The proposed workflow accepts machine-readable PDF files, scanned reports or images, and plain text. PDF extraction or Optical Character Recognition (OCR) obtains the report text, after which Natural Language Processing (NLP) preprocessing identifies medically relevant entities, values, units and findings. A Retrieval-Augmented Generation (RAG) pipeline uses embeddings and a FAISS vector index to retrieve relevant passages from a curated medical knowledge base. The retrieved context and report content are supplied to a Large Language Model (LLM) to generate a simplified explanation in English or Tamil. The system design also includes a report-aware chatbot, report history and simplified PDF export. The project is intended to support comprehension rather than diagnosis or treatment. The current project report describes illustrative outputs and qualitative behaviour; it does not establish clinical validation or quantitative accuracy. Future evaluation should measure extraction quality, factual consistency, readability and response time with suitable datasets and expert review.

KEYWORDS: Medical report simplification, Natural Language Processing, Optical Character Recognition, Retrieval-Augmented Generation, FAISS, Large Language Model, multilingual healthcare.

I. INTRODUCTION

Medical reports communicate findings from laboratory tests, radiology, pathology and prescriptions, but their clinical terminology and abbreviations can be challenging for non-specialist readers. Patients may need to wait for a consultation or search online for explanations, where information may be inconsistent or difficult to interpret. A patient-facing explanation can improve accessibility, provided that it does not alter the source report or imply an unsupported diagnosis.

Recent Natural Language Processing and Large Language Models can transform technical text into more accessible language. However, generation from a language model alone may introduce inaccurate or unsupported statements. MediSimplify AI addresses this concern through a workflow that combines extraction, NLP-based structuring and retrieval of relevant medical context before generating an explanation. The system is designed to preserve numerical values, units, findings and reference ranges, and to make the explanation available in English or Tamil.

The main contributions of the proposed design are: (1) support for PDF, image and text inputs; (2) a RAG pipeline using embeddings and FAISS retrieval; (3) bilingual English and Tamil explanations; and (4) report-specific follow-up questions, history and simplified PDF export. The system is an academic prototype and is not intended to replace professional medical advice.

II. LITERATURE SURVEY

Prior work has explored the use of large language models to simplify radiology reports for patients. Khazanchi et al. studied patient-facing explanations for lumbar spine imaging reports, illustrating the potential of LLM-based simplification while focusing on a specific report category [1]. Gupta et al. compared LLMs and prompting approaches for simplifying radiology reports for patients with cancer [4]. These studies motivate the use of readable language, while also highlighting the importance of evaluating whether clinically relevant meaning is retained.

NLP-based systems have also been proposed to identify medical terms and generate explanations. Mariammal et al. describe a medical report simplification approach using NLP-based extraction and AI-generated explanations [3]. Work on privacy-aware report interpretation highlights the importance of accessibility and responsible handling of health information [2].

Retrieval-Augmented Generation combines a language model with retrieved external passages to provide additional context during generation [7]. Sentence-BERT provides sentence embeddings useful for semantic similarity search [5], while FAISS supports efficient similarity search over vectors [6]. In MediSimplify AI, these techniques are combined with OCR, multilingual output and a report-aware chatbot in a single proposed workflow. The literature summary is based on the references listed in the project report; the bibliographic details should be verified against the original publications before submission.

III. PROPOSED SYSTEM AND METHODOLOGY

A. Input and Text Extraction: The system accepts a machine-readable PDF, a scanned PDF or image, or plain text. Embedded PDF text is extracted directly where available. OCR is applied to scanned or image-based documents to convert visible characters into machine-readable text. Image preprocessing, such as grayscale conversion and contrast adjustment, may help with noisy scans.

B. NLP Preprocessing: Extracted text is normalized to reduce excess whitespace, line-break artefacts and OCR noise. Pattern-based parsing and NLP entity recognition are used to identify relevant test names, values, units, reference ranges and free-text findings. These structured elements are retained alongside the report text so that generation can refer back to the original information.

C. Medical Knowledge Retrieval: The report content and identified entities are encoded as vector embeddings using a sentence-embedding model. A FAISS index built from a curated collection of trusted medical reference passages is searched to retrieve relevant context. The quality and currency of the knowledge base are important dependencies; retrieved material should be reviewed and maintained by suitable domain experts.

D. Grounded Explanation Generation: The original report content, structured entities and retrieved passages are supplied to an LLM with explicit instructions to use plain language and preserve the source values, units, findings and reference ranges. The generated explanation is organized as an overall summary, explanations of individual findings and definitions of technical terms. The system should avoid making diagnoses or recommending treatment.

E. Multilingual Output and Follow-up: The explanation can be generated in English or Tamil. A report-aware chatbot reuses the report context to answer clarification questions. The report-history module stores references to previously processed reports and explanations, while the PDF-export module formats a simplified summary for later reference.

F. Privacy and Safety: Medical reports may contain sensitive personal information. The design includes authenticated access, controlled report retrieval, secure API communication, limited retention of temporary extraction artefacts and a clear disclaimer. These controls must be tested in the deployed implementation; their description alone does not constitute a security audit.

IV. SYSTEM WORKFLOW

The end-to-end workflow is: (1) user uploads a report; (2) PDF extraction or OCR produces text; (3) NLP preprocessing cleans the text and identifies medical entities; (4) embeddings are generated and FAISS retrieves relevant passages from the medical knowledge base; (5) the report text and retrieved context are passed to the LLM; (6) the output is presented in English or Tamil with a safety disclaimer; and (7) the user may ask follow-up questions, revisit the report history or export a simplified PDF.

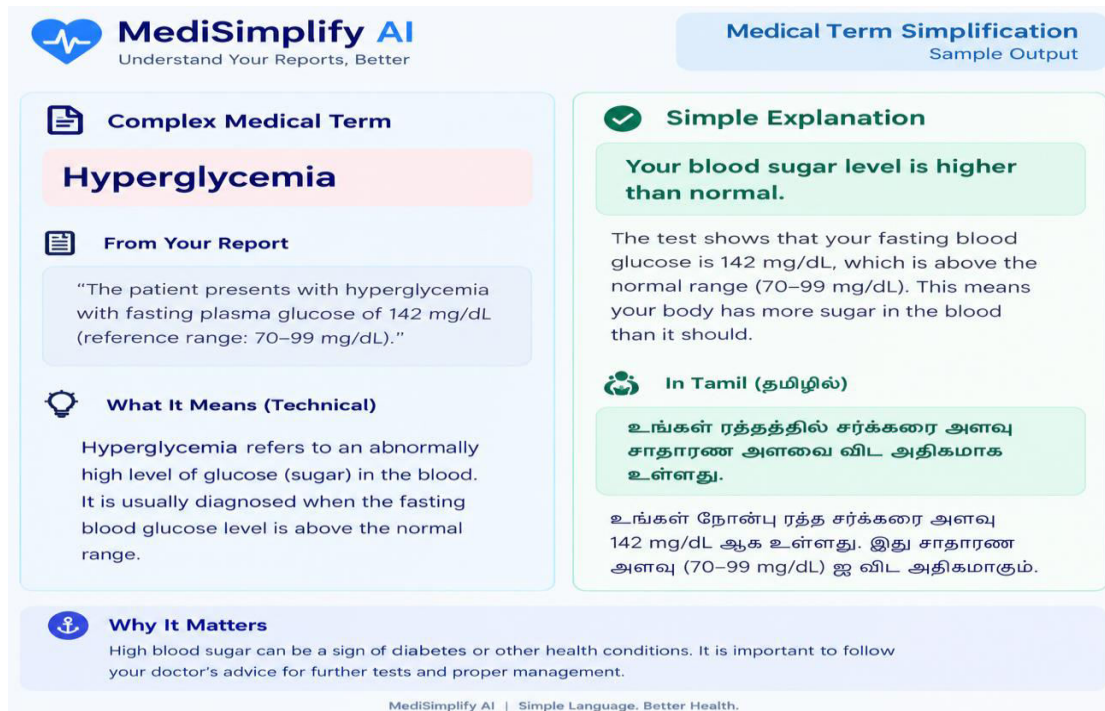
The modular design separates extraction, preprocessing, retrieval, generation and user-facing functions. This allows components to be tested individually and updated without redesigning the full pipeline. The report describes FastAPI for backend APIs, a web frontend, a relational database for user/report metadata and a vector index for retrieval.

V. RESULTS AND DISCUSSION

The project report presents illustrative output structures and qualitative discussion rather than a completed, peer-reviewed clinical validation study. Accordingly, the results in this paper are limited to the intended system behaviour and the capabilities represented by the design; no numerical accuracy, readability, latency or clinical-performance claims are made.

The expected output presents an overall summary, key findings and plain-language explanations of medical terms while keeping original values, units and reference ranges visible. The bilingual workflow is intended to communicate the same report meaning in English and Tamil. The chatbot is designed to answer report-specific clarification questions, and the history and PDF functions support later review. RAG retrieval is intended to provide contextual grounding, but it does not guarantee that every generated statement is correct.

A complete evaluation should use a suitably curated set of de-identified reports and include OCR/text-extraction accuracy, preservation of numeric values and units, factual consistency with source reports, readability, English–Tamil meaning consistency, response latency, usability and security tests. Qualified clinical reviewers should assess correctness and harmful omissions before any real-world healthcare use. These measurements remain future work and must not be reported as completed unless they are actually conducted.



MediSimplify AI
Understand Your Reports, Better

Medical Term Simplification
Sample Output

Complex Medical Term

Hyperglycemia

From Your Report

"The patient presents with hyperglycemia with fasting plasma glucose of 142 mg/dL (reference range: 70–99 mg/dL)."

What It Means (Technical)

Hyperglycemia refers to an abnormally high level of glucose (sugar) in the blood. It is usually diagnosed when the fasting blood glucose level is above the normal range.

Why It Matters

High blood sugar can be a sign of diabetes or other health conditions. It is important to follow your doctor's advice for further tests and proper management.

Simple Explanation

Your blood sugar level is higher than normal.

The test shows that your fasting blood glucose is 142 mg/dL, which is above the normal range (70–99 mg/dL). This means your body has more sugar in the blood than it should.

In Tamil (தமிழில்)

உங்கள் ரத்தத்தில் சர்க்கரை அளவு சாதாரண அளவை விட அதிகமாக உள்ளது.

உங்கள் நோன்பு ரத்த சர்க்கரை அளவு 142 mg/dL ஆக உள்ளது. இது சாதாரண அளவு (70–99 mg/dL) ஐ விட அதிகமாகும்.

MediSimplify AI | Simple Language. Better Health.

VI. CONCLUSION

MediSimplify AI proposes an integrated approach to making medical reports more understandable to patients. It combines PDF and OCR-based extraction, NLP preprocessing, retrieval using embeddings and FAISS, and LLM-based explanation generation. English and Tamil output, report-aware follow-up interaction, report history and simplified PDF export extend the system beyond a one-time summary. The design emphasizes preserving the source report and using retrieved context, while clearly positioning the application as an understanding aid rather than a diagnostic or

treatment system. Future work should prioritize systematic quantitative evaluation, expert review of clinical correctness, stronger privacy validation and improved handling of low-quality scans.

REFERENCES

- [1] R. Khazanchi et al., “Assessing the Ability of Large Language Models to Simplify Lumbar Spine Imaging Reports into Patient-Facing Text: A Comparative Study,” *Skeletal Radiology*, 2025. Verify full author list, volume, pages and DOI.
- [2] A. Mathews, A. Khan, N. T. Nairan, and M. A., “Privacy-Aware AI, Health Report Interpretation, Large Language Models, Medical Summarization, Explainable AI,” *IEEE 6th International Conference*, 2026. Verify the exact title and publication metadata.
- [3] G. Mariammal, K. Swetha, and S. Samuel Jimrys, “Medical Report Simplification System: Enhancing Healthcare Accessibility using NLP-based Extraction and AI-Driven Explanation,” *IEEE 6th International Conference on ICICV*, 2025. Verify publication details.
- [4] A. Gupta et al., “Provision of Radiology Reports Simplified with Large Language Models to Patients with Cancer,” *JCO Clinical Cancer Informatics / ASCO*, 2025. Verify final bibliographic details.
- [5] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” *Proceedings of EMNLP-IJCNLP*, 2019.
- [6] J. Johnson, M. Douze, and H. Jégou, “Billion-scale similarity search with GPUs,” *IEEE Transactions on Big Data*, 2019.
- [7] P. Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [8] Google, “Tesseract OCR Engine Documentation.” Open-source documentation. Access date and URL should be added before submission.
- [9] FastAPI, “FastAPI Documentation.” Official project documentation. Access date and URL should be added before submission.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details