Indian Oil Corporation Limited, Guwahati Refinery

Internship Project Abstract

Project Title

AI-Powered Medical Bill Verification for IOCL Employees

Organization

Indian Oil Corporation Limited (IOCL)

Submitted by: (Team: NeuroVector)

Team Members: Aviraj Roy(C), Biswadeep Das

Abstract

Purpose and Scope:

This project enables automated processing of medical reimbursement claims submitted by IOCL employees. Scanned bills and receipts are ingested by the system, which extracts key information (patient name and ID, hospital name, dates of service, procedures, medications, charges, etc.) using OCR and organizes it into structured data. The extracted data is then compared against IOCL's database of pre-approved "tie-up" hospital rates and covered services. Only IOCL employee claims are in scope, ensuring that each claim is vetted internally for compliance with negotiated rates and policy coverage.

Technology Stack:

The solution combines proven OCR engines with state-of-the-art LLMs. For OCR we will evaluate engines such as Tesseract(open-source), PaddleOCR(used in production systems), Donut(used in invoice parsing); these tools achieve high recognition accuracy on healthcare forms. For the comparison stage we will use a large language model (such as GPT-4 or similar) to interpret the structured data. The LLM will be supported by a retrieval mechanism (RAG) that queries IOCL's reference dataset of contracted rates and services. In this way, the model can "understand" each bill item in context and check it against the relevant agreements.

Methodology

1. OCR Extraction and Data Structuring:

Each submitted bill (PDF or image) is processed by the OCR engine, which recognizes text and fills predefined fields (patient name & ID, hospital, doctor, test or procedure names, medication names and dosages, billed amounts, etc.). This transforms unstructured images into clean, grouped records. The extracted fields are validated and organized into a standardized format.

2. LLM-Based Verification: The structured claim data is fed to the LLM-powered analysis engine. Using natural-language understanding, the LLM cross-references each charge and item against IOCL's approved tie-up rates and hospital service agreements. The model flags any mismatches or unusual items, for example, charges that exceed the negotiated rate or items not covered by the agreement. In practice, the LLM acts as an intelligent auditor: querying the reference database and generating explanations for any discrepancies (e.g. "Room charges billed at X per day vs. approved Y")

Output Format

The system will produce a clear summary report or interactive web dashboard for claims officers. For each processed bill, the output will list the extracted fields side-by-side with the corresponding reference rates, and highlight exceptions.

For example, an HTML/PDF report might display the patient's services and costs in a table, with any overcharges or non-covered items flagged in red. This structured output can be reviewed online or downloaded; it could also integrate with IOCL's existing claim processing workflow. By turning opaque bills into annotated data tables, the system gives reviewers immediate visibility into each claim.

Benefits to IOCL

1. Cost Transparency and Control: Automated cross-checking ensures that every billed item is explicitly matched against the agreed corporate rate. This eliminates hidden mark-ups or coding errors and makes costs fully transparent. Any deviation from contract terms is immediately visible, so IOCL can enforce negotiated pricing.

2. Fraud and Overcharge Detection: The AI engine will promptly flag suspicious charges (duplicate services, inflated quantities, unauthorized tests, etc.). LLM-based analysis has been shown to identify such billing anomalies with very high accuracy. By catching these issues automatically, the system helps prevent fraud, waste, and unnecessary payments.

3. Operational Efficiency: By replacing manual data entry and laborious rate lookups, the automated pipeline dramatically speeds up claim processing. Instead of clerks retyping and verifying each bill, the OCR+LLM solution handles routine cases instantaneously. In practice, organizations report that AI-powered OCR can cut processing time by roughly half and reduce error rates significantly. This frees IOCL staff to focus only on exceptions, reducing backlog and turnaround time.

Overall, the project will improve accuracy while enabling faster reimbursements and better allocation of administrative resources

In summary, this in-house system will turn employee medical bills into actionable data: extracting all key fields automatically, comparing them against IOCL's contracts, and producing concise reports for reviewers. The result is greater cost transparency, rigorous enforcement of agreed rates, and more efficient claims operations, helping IOCL contain costs and detect any irregular claims early.