

Data Extraction for Faster Document Processing

Streamlining medical diagnostic workflows by automating complex data extraction from diverse lab reports using AI-driven document processing.

Overview

Developed an AI-based data extraction engine leveraging OCR, Computer Vision, and Natural Language Processing (NLP) to digitize scanned PDF reports.

- Integrated a pluggable domain-specific layer and specialized algorithms to accurately capture handwritten text and patient details.
- Achieved significant operational efficiency by automating data flows into third-party applications, reducing manual intervention and error rates.



Client Profile

The client is a premier medical diagnostics company delivering a wide range of laboratory testing services to healthcare providers.

Challenges: Overcoming Manual Bottlenecks

High variability in lab report layouts made traditional template-based parsing impossible.

- Handwritten Overlaps: Critical test data and notes were often handwritten, sometimes overlapping with printed text.
- Scalability Issues: Manual data entry was resource-intensive, slow, and prone to human error, hindering large-scale processing.

QBurst Solution: Intelligent AI Extraction

We implemented a robust Intelligent Document Processing (IDP) solution that treats every report as a unique data set. By combining Computer Vision with OpenAI's language capabilities, the system understands the context of the medical data it extracts.

- **Advanced OCR & HTR:** Used image-based models to handle skewed scans and sophisticated algorithms to recognize handwritten text at varying angles.
- **Pluggable Domain Layer:** A flexible architecture allows for custom extraction logic tailored to specific healthcare domains or report types.
- **Precision Logic:** Built-in error rejection mechanisms identify and filter out low-confidence outputs to maintain medical-grade data integrity.

Technical Highlights

- **Multi-source data extraction:** Capable of extracting personal details, test information, and handwritten data from scanned PDF lab reports.
- **Robust image processing:** Utilizes image-based models to adjust for skewed images and minor orientation issues, ensuring accurate data extraction.
- **Domain-specific customization:** Offers a pluggable domain-specific layer allowing customization for different domains, enhancing accuracy and relevancy.
- **Handwritten text recognition:** Specialized algorithms to extract handwritten data from scanned documents, handling varying angles, formats, and overlaps with printed text.
- **Predefined data coordinates support:** Ability to define and extract data from specific areas within documents, ensuring higher accuracy in extracting targeted information.
- **Adaptability and customization:** Customizable to handle multiple forms and domains, ensuring flexibility in data extraction processes.
- **Error rejection logic:** Incorporates logic to reject inaccurate or less reliable outputs, improving overall data accuracy and integrity.

Impact

The IDP solution improved efficiency and resulted in a drastic reduction in manual effort required for analyzing and entering lab report data.

- **Enhanced Accuracy:** Domain-specific logic and automated validation significantly increased the reliability of patient records.
- **Cost Savings:** Repetitive clerical tasks were eliminated, allowing the workforce to focus on high-value diagnostic tasks.
- **Versatile Scalability:** The customizable framework easily adapts to new forms and healthcare domains as the company grows.