

# **TECHNICAL DETAILS:** DROP & SORT - SECURE AI-DRIVEN DOCUMENT PROCESSING PLATFORM

DP & SORT - SECURE AI-DRIVEN DOCUMENT PROCESSING PLATFORM	l
Core Capabilities	
2. Document Processing Worknow	
Phase One: Initial Mechanical Splitting	
Phase Two: Intelligent Splitting	
Phase Four: Cleaning and Trimming	
Phase Five: Final Merging and Indexing	
Submissions Dashboard	
Notifications	
. Technology Stack	
Frontend	
Backend	6
Azure Services	6
Notifications	7
Database	7
Additional Libraries	8
Notification Workflow Example	
Example Batch Status Notification	8
Why Azure Services Were Chosen	٤
Security Features	9
4.1 Document Security	
4.2 Role-Based Access Control (RBAC)	
4.3 Auditing	ç
5. Document Processing Highlights	ç
6. Compliance and Data Residency	
Compliance	
Data Residency	
7. Summary	



## 1. Overview of the Platform

Drop & Sort platform is a secure, scalable, and Al-driven solution for processing **medical records** used by **insurance companies**. Its primary purpose is to process, analyze, organize and securely manage sensitive medical documents, extracting valuable metadata and generating comprehensive summaries to streamline workflows.

## **Core Capabilities**

- 1. Automated Document Analysis:
  - Separate multi-document files into single logical documents.
  - Extract metadata such as **Date of Service**, **Document Type**, **Title**, **Facility**, **Provider**, and **Specialty**.
  - Analyze documents as a whole to extract additional relevant information and generate summaries.
  - Determine and remove redundant or irrelevant information from processed documents.
- 2. User Management:
  - Registration, role-based access control, and admin approval.
  - Role hierarchy: Admins, Branch Managers, and Clients.
- 3. Compliance:
  - Ensures adherence to healthcare privacy standards such as HICFFA, SOC II, PHI, ISO, WHMIS, and PIPEDA.

## 2. Document Processing Workflow

Each file within a batch undergoes a multi-step processing workflow, where files are analyzed, split, cleaned, and merged into a final comprehensive report. The workflow consists of five distinct **Phases**, ensuring efficient processing and compliance with data standards.

## Phase One: Initial Mechanical Splitting

#### 1. Purpose:

- Handle large files exceeding **80 pages** by splitting them into manageable pieces.
- This step is necessary because AI models, such as the one used for document processing, have limitations regarding the physical size of documents they can handle—typically capped at around 25 MB or 100 pages. After careful analysis of these constraints and to ensure smooth processing, we determined that splitting larger documents into 80-page chunks strikes an optimal balance between efficiency, quality, precision and compatibility with the AI's processing capabilities.



#### 2. How It Works:

- The system uses a mechanical splitting algorithm to divide oversized files into 80-page segments.
- Split files are saved as individual pieces in **Azure Blob Storage**.
- These files are flagged for further intelligent processing in **Phase Two**.
- Original page numbering is preserved for reassembly in the final report.

### Phase Two: Intelligent Splitting

- 1. Purpose:
  - Split multi-document files into logical, single-document files for individual processing.

#### 2. How It Works:

- Using **AI-powered content analysis**, the system identifies logical boundaries such as:
  - Headers or footers.
  - Separators and table of contents.
  - Document structures (e.g., page breaks, sections).
- Split documents are stored as independent files in Azure Blob Storage.

#### Phase Three: Individual Document Analysis

#### 1. Metadata Extraction:

- Purpose:
  - Extract essential metadata to classify and index documents accurately.
- Metadata Fields:
  - **Date of Service**: Extracted from document headers or content.
  - Document Type: Categorized (e.g., therapy notes, billing, discharge summaries).
  - **Document Title**: Based on document structure or semantic cues.
  - **Facility**: Name of the healthcare facility.
  - **Provider**: Name of the doctor or medical professional.
  - **Specialty**: E.g., physical therapy, cardiology.
- How It Works:
  - The OpenAl API processes the text, extracts metadata, and saves it in Azure Cosmos DB.
- 2. Content Analysis and Summary Generation:



- Purpose:
  - Analyze full content for relevant information and generate a concise summary.
- Key Features:
  - Identify critical sections such as diagnoses, treatments, and recommendations.
  - Compile a meaningful summary for insurance claim assessments.
- How It Works:
  - **OpenAl language models** analyze document content.
  - Summaries and additional extracted information are stored in the database.

#### **Phase Four: Cleaning and Trimming**

- 1. Repetitive/Irrelevant Content Detection:
  - Purpose:
    - Identify and remove redundant or non-essential content, such as:
      - Duplicate sections.
      - Signatures or confidentiality notices.
      - Repeated tables or disclaimers.
  - How It Works:
    - The system uses Al-powered prompts tailored to the document type (e.g., bill, claim, health note) to detect repetitive sections.

#### 2. Mechanical Removal:

- Purpose:
  - Physically remove detected sections to reduce file size and improve document clarity.
- How It Works:
  - Unnecessary sections are removed while preserving original page numbering.
  - Modifications are logged in a CSV file, detailing:
    - Removed sections.
    - Original page locations.
- Outputs:
  - Cleaned PDF files are stored in **Azure Blob Storage**.
  - Modification logs are saved for audit purposes.

### **Phase Five: Final Merging and Indexing**

- 1. Merge Split Documents:
  - Purpose:



- Combine all single documents from a batch into a final report, n pagination.
- How It Works:
  - Split and cleaned files are merged into a single PDF report using libraries like PDF-Lib.
  - The order is based on the original file and document sequence.

#### 2. Generate Index Spreadsheet:

- Purpose:
  - Create a searchable, structured metadata index for the report.
- How It Works:
  - An Excel file is generated with:
    - Metadata (e.g., document type, title, provider).
    - Page ranges and corresponding file details.

### **Submissions Dashboard**

- 1. User Access:
  - Clients:
    - View their uploaded batches and track processing status.
  - Branch Managers:
    - Oversee submissions for their branch and access related reports.
  - Admins:
    - Monitor all submissions across the platform.

#### 2. Dashboard Information:

- Batch Details:
  - Batch ID, submission date, number of documents.
- Status:
  - Pending, Accepted, Processing, Completed.
- File Links:
  - Links to download the final report and the index file.

#### **Notifications**

- Triggered Events:
  - Batch acceptance after format validation.
  - Status updates during Phases (e.g., Accepted, Processing, Completed).
  - Notifications sent via Azure Communication Services (ACS).



## 3. Technology Stack

The platform leverages a robust combination of tools, libraries, and Azure services to ensure secure, scalable, and efficient processing of medical records.

## Frontend

Technology	Purpose	
React	Secure and scalable frontend development.	
Material UI	Component library for modern UI/UX.	

## Backend

Technology	Purpose	
Node.js/Express	Secure API development.	
Axios	For handling API requests.	
PDF-Lib	Parsing, splitting, and merging documents.	
OpenAl API	Al-powered document borders determination, metadata extraction, summary generation, irrele content search.	
jsonschema	Input validation for secure API endpoints.	

## **Azure Services**

Azure Service	Purpose
Azure Blob Storage	Stores raw, processed, and final reports with encryption at rest.
Azure Service Bus	Task queue for handling batch processing and document separation tasks in parallel.
Azure Durable Functions	Orchestrates workflows for document processing, role approval, and notifications.



Azure Cosmos DB	NoSQL database for storing metadata, user roles, and batch statuses.	
Azure SQL Database (Option	For structured storage and complex querying needs (e.g., analytics).	
Azure Monitor	Tracks platform performance, logs API requests, and monitors user actions for auditing.	
Azure Application Insights	Provides real-time insights into application performance and error detection.	
Azure Security Center	Irity Center Monitors and protects the infrastructure against security threats.	
Azure Key Vault	Manages secrets, API keys, and connection strings securely.	
Azure Communication Serv (ACS)	Sends batch status notifications via email, SMS, or other channels.	

## Notifications

Service	Purpose	
Azure Communication Servi (ACS)	Notify clients when their batch status changes (e.g., Pending, Processing, Completed, Rejec	
Integration Features	Email, SMS, or push notifications can be configured based on user preferences.	

## Database

Technology	Technology Purpose	
Azure Cosmos DB	Metadata storage, user management, and auditing.	
SQL Database (Optional)	ase (Optional) For structured query storage if needed.	



## Additional Libraries

Library	Purpose	
jsonwebtoken	JWT-based authentication.	
bcrypt	Secure password hashing.	
Winston	Logging API activities for auditing.	

### **Notification Workflow Example**

#### 1. Trigger Notification:

- Notifications are triggered by changes in batch status (e.g., a batch moves from Processing to Completed).
- 2. Durable Function Notification Workflow:
  - The Azure Durable Function monitors status updates in Cosmos DB or a SQL trigger.
  - When an update occurs, the function sends a notification using **Azure Communication Services (ACS)**.



- 2. Security:
  - Azure Security Center and Key Vault protect sensitive data and credentials.
- 3. Flexibility:
  - Azure Communication Services provides multiple notification channels tailored to client needs.

#### 4. Compliance:

 Azure's services comply with SOC II, HIPAA, PIPEDA, and other healthcare regulations, ensuring seamless alignment with project requirements.



## 4. Security Features

## **4.1 Document Security**

- Encryption:
  - **At Rest**: Files are encrypted using AES-256.
  - In Transit: TLS 1.2+ ensures secure communication between components.
- Access Control:
  - Files are accessible only via time-limited SAS tokens.

## 4.2 Role-Based Access Control (RBAC)

- Clients: Can upload and view their own processed files.
- Branch Managers: Can access all submissions from their branch.
- Admins: Full access across the platform.

### 4.3 Auditing

• Every user action (upload, download, approval) is logged in **Azure Monitor** for auditing and compliance.

## **5. Document Processing Highlights**

Phase	Stage	Technology/Approach	Output
1	Mechanical Separation (for large files)	PDF-Lib	80-page chunks to comply with Al processing limitations
2	Intelligent Separation	PDF-Lib and AI-based analysis	Single logical documents.
3	Metadata Extraction	OpenAI API	Date, type, title, facility, provider, specialty.
	Content Analysis	OpenAl API	Additional details and summaries.
4	Cleaning and Trimming	OpenAI API and PDF-Lib	Lighter file containing only relevant data
5	Report Generation	PDF-Lib	Merged PDF report and Excel + CSV index files.



## 6. Compliance and Data Residency

## Compliance

The platform is compliant with:

- **HICFFA**: Ensures secure handling of health information.
- SOC II: Operational security and data integrity.
- **PHI**: Protects U.S. healthcare data (HIPAA-compliant).
- **PIPEDA**: Adheres to Canadian privacy laws.
- ISO 27001: Information security management.

### **Data Residency**

• All sensitive data is processed and stored in **Azure's Canadian data centers**, ensuring compliance with Canadian data sovereignty laws.

## 7. Summary

The platform combines AI-driven processing, secure infrastructure, and strict compliance measures to deliver a robust solution for processing medical records. Its multi-stage document processing workflow ensures accurate data extraction and meaningful summaries, enabling insurance companies to streamline their workflows while adhering to industry regulations.