

Introduction to
Robotic Process
Automation

Handwritten Insurance form Digitalization

220701025

Amrutha B J

Jinu Sophia

Designation and Computer Science &
Engineering



RAJALAKSHMI
ENGINEERING COLLEGE

Abstract

- The Handwritten insurance Form Digitalization solution in UiPath utilizes cutting-edge Optical Character Recognition (OCR) technology to convert handwritten loan forms into structured digital data. This automation extracts critical information such as borrower names, addresses, and loan details, seamlessly populating it into Excel sheets or databases. By eliminating the need for manual data entry, the solution saves significant time and effort while reducing human errors, ensuring high accuracy and efficiency. Designed for financial institutions, this tool streamlines the loan processing workflow and enhances data management capabilities.

Need for the Proposed System

- **Efficient Data Processing:** Automates the extraction of data from handwritten insurance forms, saving time compared to manual entry.
- **Error Reduction:** Minimizes human errors commonly associated with manual transcription.
- **Enhanced Accuracy:** Ensures reliable data extraction using advanced OCR technology.
- **Streamlined Workflow:** Speeds up insurance processing by digitizing critical information seamlessly.
- **Cost-Effective:** Reduces labor costs by eliminating repetitive manual tasks.
- **Improved Data Management:** Converts unstructured handwritten forms into structured formats for easier storage and retrieval.

Advantages of the Proposed System

1. **Time Efficiency:** Automates data extraction, drastically reducing the time required to process handwritten forms.
2. **Improved Accuracy:** Minimizes manual errors through precise OCR technology, ensuring reliable data entry.
3. **Cost Savings:** Lowers operational costs by eliminating repetitive manual tasks and reducing labor dependency.
4. **Scalability:** Easily handles large volumes of loan applications, supporting business growth without additional resources.
5. **Faster Decision-Making:** Provides quick access to digitized and structured data, expediting the insurance approval process.

Literature Survey

Paper 1: "Handwritten Text Recognition Using Machine Learning"

Advantages:

- Utilizes machine learning for improved accuracy in handwriting recognition.
- Can handle diverse handwriting styles.

Disadvantages:

- Requires extensive training data for high performance.
- May struggle with highly distorted or noisy handwriting.

Paper 2: "A Robust Handwritten Text Recognition System Based on Convolutional Neural Networks"

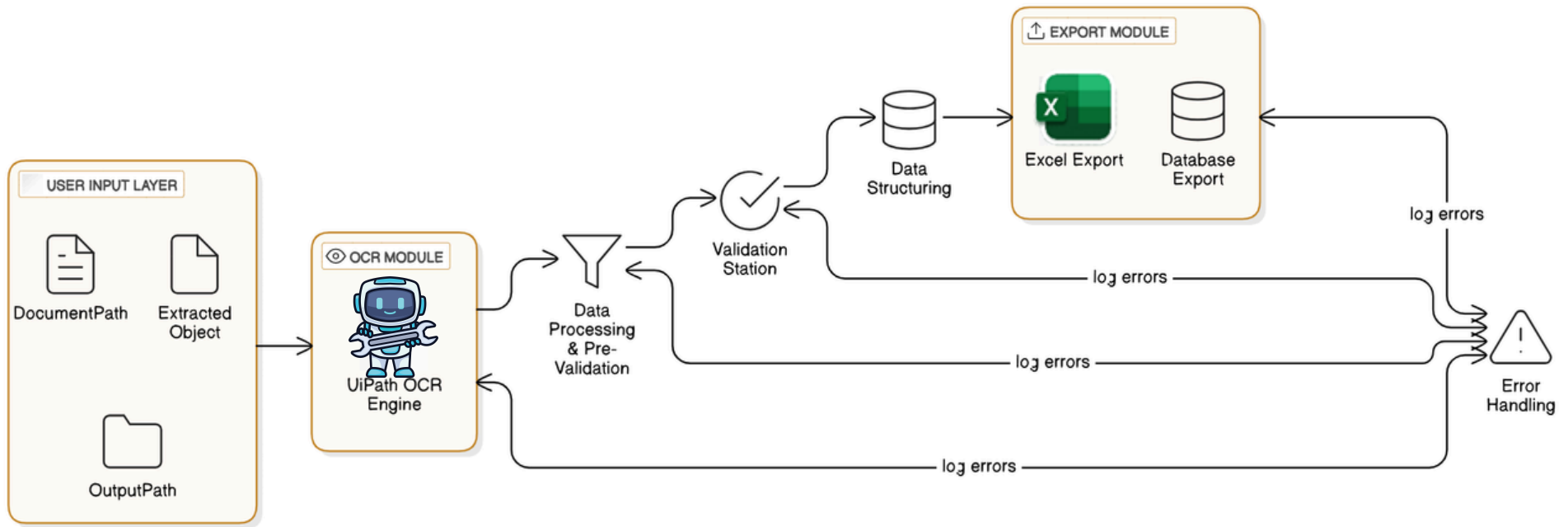
Advantages:

- Leverages CNNs for enhanced robustness and accuracy.
- Suitable for large-scale datasets.
- Disadvantages:
- High computational cost.
- Requires large annotated datasets for training.

Main Objective

- **Automate Data Extraction:** Utilize OCR technology to extract handwritten information from insurance forms.
- **Transform Data:** Convert unstructured handwritten data into structured digital formats.
- **Enhance Efficiency:** Streamline loan processing workflows and reduce processing time.
- **Minimize Errors:** Ensure accuracy in data extraction and eliminate manual transcription errors.
- **Improve Scalability:** Enable the system to handle large volumes of loan applications efficiently.

Architecture



System Requirements

Hardware Requirements

Processor: Intel Core i5 or higher.

RAM: Minimum 8 GB (16 GB recommended).

Storage: 256 GB SSD or more.

Input Devices: Scanner or camera for form capture.

Network: Stable internet connection.

Software Requirements

OS: Windows 10/11, macOS, or Linux (64-bit).

UiPath Studio: For automation workflows.

OCR Tools: Tesseract, Google Vision API.

Spreadsheet Tools: Microsoft Excel or Google Sheets.

Functional Description

Module 1: Data Capture and OCR Processing

Main Process:

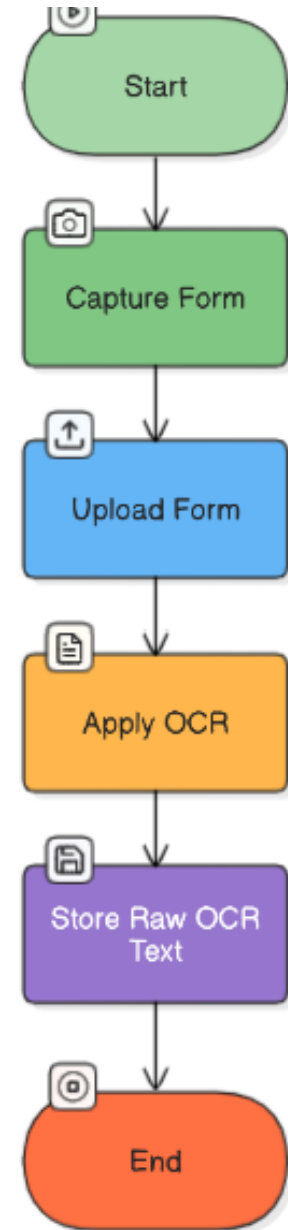
This module involves capturing the handwritten loan form and processing it through OCR.

1. Data Capture

- Capture the loan form via scanning or camera.
- Upload the form for OCR processing.

2. OCR Processing

- Apply OCR to convert handwritten text to machine-readable text.
- Store raw OCR text.



Functional Description

Module 2: Data Structuring and Email Notification

Main Process:

This module formats and stores the extracted data and sends email notifications to stakeholders.

1. Structured Data Output

- Format and structure the extracted data (e.g., names, addresses, loan amounts).
- Store data in a database or Excel.

2. Email Notification

- Compose an email with the extracted loan data.
- Send email to stakeholders (e.g., loan officers, applicants)

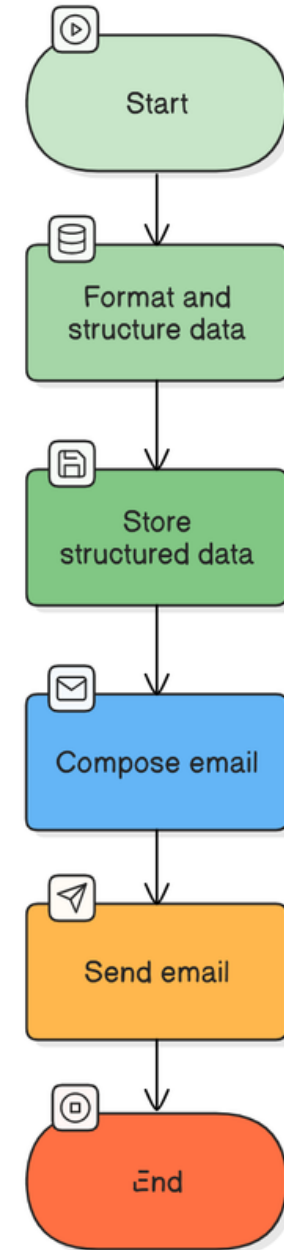


Table Design



Process Design

Main Process

This process captures the handwritten loan form and sends email notifications with processed data.

Sub-Processes:

1. Data Capture

- Capture the loan form via scanning or camera.
- Upload the form for OCR processing.

2. OCR Processing

- Apply OCR to convert handwritten text to machine-readable text.
- Store raw OCR text.

3. Structured Data Output

- Format and structure the extracted data (e.g., names, addresses).
- Store data in a database or Excel.

4. Email Notification

- Compose an email with extracted data.
- Send email to stakeholders (e.g., loan officers, applicants).

Implementation

Module 1: Data Capture and OCR Processing

RELIGARE Health Insurance
Values that bind

Ab Health Hamesha

Claim Form - 'GROUP CARE'
Part A

1. To be filled in by the Insured.
2. The issue of this Form is not to be taken as an admission of liability.
3. To be filled in block letters.

Section A - Details of Primary Insured

a) Policy No. : ABCDEFGHIJKLMNOP12345678901010ABXZ

b) SL No/Certificate No: 87654321342 c) Company/TPA ID No: xy2AGC1234

d) Name : Anand Kanakajiri
(Surname) (First Name) (Middle Name)

e) Address : 1209 35th B CROSS 4th GLOCK
JAYANAGAR

RELIGARE Health Insurance
Values that bind

Ab Health Hamesha

Claim Form - 'GROUP CARE'
Part A

1. To be filled in by the Insured.
2. The issue of this Form is not to be taken as an admission of liability.
3. To be filled in block letters.

Section A - Details of Primary Insured

a) Policy No. : ABCDEFGHIJKLMNOP12345678901010ABXZ

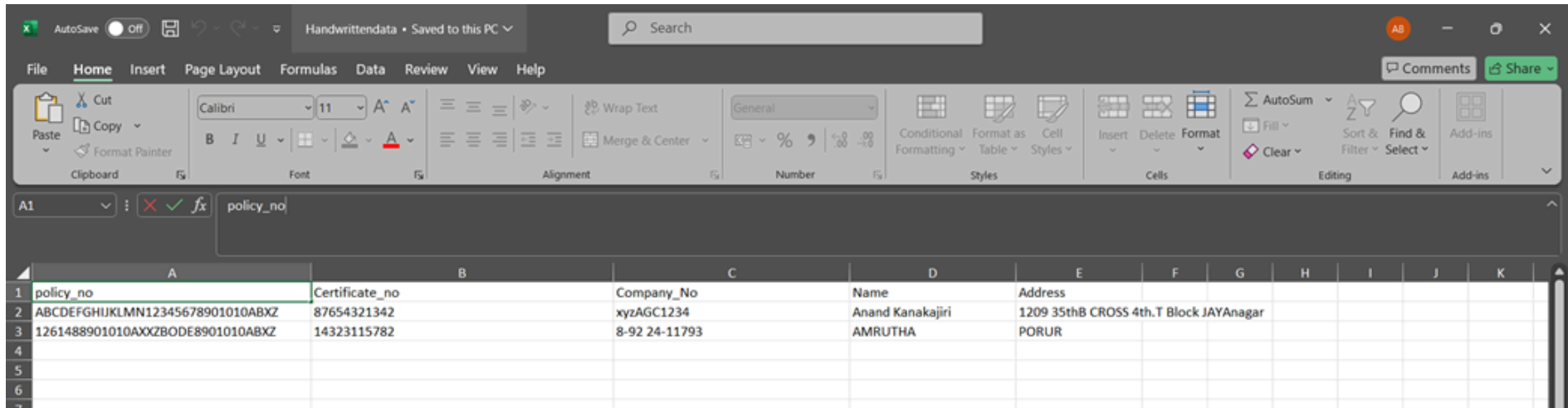
b) SL No/Certificate No: 87654321342 c) Company/TPA ID No: xy2AGC1234

d) Name : Anand Kanakajiri
(Surname) (First Name) (Middle Name)

e) Address : 1209 35th B CROSS 4th GLOCK
JAYANAGAR
City: Bangalore

Implementation

Module 2: Data Structuring and Email Notification



The screenshot shows the Microsoft Excel interface with a data table. The table has the following columns and data:

policy_no	Certificate_no	Company_No	Name	Address
ABCDEF GHIJ KLMN12345678901010ABXZ	87654321342	xyzAGC1234	Anand Kanakajiri	1209 35thB CROSS 4th,T Block JAYAnagar
1261488901010AXXZBODE8901010ABXZ	14323115782	8-92 24-11793	AMRUTHA	PORUR

Your Handwritten Data Extraction Results Are Ready!

Dear User,

We are excited to inform you that your handwritten loan document has been successfully processed! 🎉

Key Data Extracted:

- Loan Amount
- Interest Rate
- Loan Type
- Applicant Name

📎 The extracted data is attached in an Excel file for your easy reference.

[Download and Review Your Data](#)

If you have any questions or need further assistance, feel free to reply to this email. We're here to help!

Best regards,

Your Automation System

One attachment • Scanned by Gmail



Testing

Data Capture & OCR Processing

- Test image/scan quality and file compatibility (JPG, PNG, PDF).
- Validate OCR accuracy with various handwriting styles and formats.
- Ensure proper extraction of fields like names, addresses, and loan amounts.

Data Structuring & Email Notification

- Verify extracted data is correctly structured and stored in a database/Excel.
- Test email composition accuracy and ensure successful delivery to stakeholders.

Conclusions

The Handwritten insurance Form Digitalization project successfully automates the process of extracting and structuring data from handwritten insurance forms. By utilizing Optical Character Recognition (OCR), the project eliminates the need for manual data entry, significantly reducing time and human error.

Through two key modules — Data Capture and OCR Processing and Data Structuring and Email Notification — the system captures, processes, formats, and stores extracted data efficiently. The integration of email notifications ensures timely communication with relevant stakeholders, improving overall workflow and data accessibility.

In conclusion, this project enhances operational efficiency by streamlining the data extraction process, reducing errors, and enabling faster decision-making, making it a valuable solution for organizations handling handwritten loan forms.

Future Enhancement

Advanced OCR Accuracy with AI Integration:

- Integrating advanced AI models and deep learning techniques to improve OCR accuracy, especially for complex handwriting, would enhance the system's ability to extract data from diverse handwriting styles and unclear forms.

Real-time Data Extraction and Processing:

- Implementing real-time data extraction capabilities would allow users to upload loan forms and instantly receive processed data, improving the overall speed and efficiency of the system.

IEEE Paper

- "Handwritten Text Recognition Using Machine Learning" – This paper focuses on using machine learning techniques for handwritten text recognition, particularly exploring new methods for improving accuracy in OCR systems [click here](#)
- "A Robust Handwritten Text Recognition System Based on Convolutional Neural Networks" – This paper discusses the application of convolutional neural networks (CNNs) in enhancing handwritten text recognition, with a focus on improving system robustness.
<https://ieeexplore.ieee.org/document/10303304>

References

1. "Pattern Recognition" – This journal focuses on research in pattern recognition, including handwriting recognition using machine learning and deep learning algorithms.
2. "International Journal of Document Analysis and Recognition (IJDAR)" – A journal dedicated to research in document analysis, including OCR and handwritten text recognition.
3. "Journal of Machine Learning Research (JMLR)" – Publishes papers on machine learning algorithms that can be applied to handwritten text recognition tasks.

Thank
You