DIGITIZING INSURANCE APPLICATIONS: HANDWRITTEN DATA EXTRACTION

A PROJECT REPORT

Submitted by

AMRUTHA B J (2116220701025)

in partial fulfillment for the course

OAI1903 - INTRODUCTION TO ROBOTIC PROCESS AUTOMATION

for the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

RAJALAKSHMI NAGAR

THANDALAM

CHENNAI – 602 105

NOVEMBER 2024

BONAFIDE CERTIFICATE

Certified that this project report **"Handwritten Insurance Document Digitalization"** is the bonafide work of **"AMRUTHA B J (220701025)"** who carried out the project work for the subject OAI1903 - Introduction to Robotic Process Automation under my supervision.

Mrs. J. Jinu Sophia

SUPERVISOR

Assistant Professor (SG)

Department of

Computer Science and Engineering

Rajalakshmi Engineering College

Rajalakshmi Nagar

Thandalam

Chennai - 602105

Submitted to Project and Viva Voce Examination for the subject OAI1903 - Introduction to Robotic Process Automation held on_____.

Internal Examiner **External Examiner**

ABSTRACT

The Handwritten Document Digitalization in UiPath project leverages Optical Character Recognition (OCR) to automate the extraction of data from handwritten insurance documents. By using UiPath's robotic process automation (RPA) capabilities, it significantly enhances operational efficiency by converting handwritten data such as policyholder names, coverage details, claims, and premium information into structured digital formats.

This extracted information is then seamlessly transferred to Excel sheets or databases for further processing, ensuring accurate and consistent data management. By automating the process, this project reduces manual workload, improves accuracy, and speeds up processing times, making it an invaluable tool for insurance companies aiming to optimize their document management, reduce human error, and improve overall productivity.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. Meganathan, B.E, F.L.E.,** our Vice Chairman **Mr. Abhay Shankar Meganathan, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) Thangam Meganathan, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution

Our sincere thanks to **Dr. S.N. Murugesan, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time We express our sincere thanks to **Dr. P. Kumar, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides, **Mrs. J. Jinu Sophia, M.E., (Ph.D.),** Assistant Professor (SG), Department of Computer Science and Engineering. Rajalakshmi Engineering College for her valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinators, **Dr. N. Durai Murugan, M.E., Ph.D.,** Associate Professor, and **Mr. B. Bhuvaneswaran, M.E.**. Assistant Professor (SG), Department of Computer Science and Engineering for their useful tips during our review to build our project

Amrutha B J (2116220701025)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATION	ACCRONYM
RPA	Robotic Process Automation
AI	Artificial Intelligence
API	Application Programming Interface
OCR	Optical Character Recognition

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

The process of handling insurance claims often involves dealing with large volumes of handwritten documents, where critical data such as policyholder details, coverage, and claims need to be extracted accurately and efficiently. Traditionally, this task requires significant manual effort to read, interpret, and input data from handwritten forms into digital systems, which is time-consuming and prone to human errors, resulting in delays and inaccuracies. As insurance companies continue to scale, the need for an efficient, reliable, and automated solution becomes increasingly important.

The Handwritten to Text Form Extraction in UiPath project aims to address these challenges by automating the extraction of handwritten data from insurance documents. Using Optical Character Recognition (OCR) technology combined with UiPath's Robotic Process Automation (RPA) platform, this project allows for the seamless conversion of handwritten text into structured, machine-readable data. The system extracts essential information such as policyholder names, coverage details, claims data, and premium amounts with high accuracy.

Once extracted, the data is automatically populated into digital formats like Excel sheets or databases, eliminating manual data entry. This automation streamlines the insurance claim process, reduces processing times, and ensures that key information is accurately captured, minimizing the risk of human error.

The solution not only enhances efficiency but also reduces operational costs by minimizing manual effort. By automating repetitive tasks, insurance companies can redirect resources toward more value-added activities, such as customer service or advanced data analytics. Furthermore, it ensures faster processing, improving customer satisfaction and overall business productivity.

This project represents a transformative step in modernizing how insurance companies manage handwritten documents, providing a faster, more accurate, and scalable solution to longstanding industry challenges.

1.2 OBJECTIVE

The objective of the Handwritten to Text Form Extraction in UiPath project is to automate the extraction of handwritten data from loan forms using Optical Character Recognition (OCR) technology and UiPath's Robotic Process Automation (RPA) tools. This solution aims to improve the accuracy and efficiency of loan processing by eliminating manual data entry, reducing errors, and converting the extracted information into structured formats like Excel sheets or databases. By streamlining data extraction and processing, the project seeks to accelerate loan approval times, reduce operational costs, and enhance overall customer satisfaction.

1.3 EXISTING SYSTEM

Currently, most insurance companies rely on manual processes to extract data from handwritten insurance forms. These forms are typically reviewed and transcribed by staff, a time-consuming and error-prone task. The existing system often involves scanning handwritten documents followed by manual data entry into digital systems, such as databases or spreadsheets. This approach leads to delays in processing claims, increases the likelihood of human errors like incorrect data entry or missed information, and makes handling large volumes of claims cumbersome. Basic OCR tools may be used but often require significant human oversight for accuracy.

1.4 PROPOSED SYSTEM

The Handwritten to Text Form Extraction Bot revolutionizes the insurance document processing workflow by automating the extraction of handwritten data using UiPath's Robotic Process Automation (RPA) and Optical Character Recognition (OCR) technology. The bot accurately processes insurance forms, extracting essential details like policyholder names, coverage types, claim amounts, and dates. This data is converted into structured, machine-readable formats such as Excel files or databases,

ensuring efficient and error-free processing. By automating these tasks, the bot reduces operational costs, accelerates claim processing times, and improves overall accuracy, providing a scalable solution for the insurance industry.

CHAPTER 2

LITERATURE REVIEW

2.1 Survey on Robotic Process Automation (RPA) in Insurance Document Processing

[1] Robotic Process Automation (RPA) has become a vital tool in the insurance industry, helping companies automate repetitive and time-intensive tasks, which streamlines operations and increases overall efficiency. RPA enables the automated extraction and processing of data from insurance documents such as claims, policies, and customer forms. By digitizing these processes, RPA reduces the time and cost associated with manual data entry, ensuring faster and more accurate results.

[2] In addition, RPA technology, often combined with Artificial Intelligence (AI) and Optical Character Recognition (OCR), is capable of interpreting complex data from unstructured sources. This integration allows RPA to handle large volumes of documents with precision, significantly improving accuracy and reducing human errors. As AI models continue to advance, the ability to interpret diverse handwriting styles and varying document formats will further enhance the system's capabilities.

[3] The insurance sector has already seen improvements in document management, where RPA has been applied to claims processing, policy underwriting, and customer service. For instance, RPA has automated the extraction of customer information from insurance claims, reducing both processing times and errors. This trend shows the growing reliance on RPA for tasks that require high accuracy, scalability, and efficiency, making it a pivotal technology for future-proofing the insurance industry's operations.

2.2 Survey on Digitalizing Handwritten Insurance Forms:

In the context of insurance document processing, RPA has demonstrated significant potential by automating the extraction, validation, and reporting of data, leading to reduced manual labor and fewer errors. Automating handwritten insurance forms has become a major focus, aiming to minimize data entry and enhance processing speed. Technologies such as Optical Character Recognition (OCR) and RPA are increasingly employed to extract data from handwritten insurance documents. However, challenges like variations in handwriting, low-quality scans, and complex document structures still hinder system accuracy. Below is a summary of research papers in this field:

[3] A study assessed OCR performance in extracting data from handwritten insurance forms, finding that OCR systems performed well on neatly written documents but struggled with poor-quality handwriting and scans. The research emphasized the need for manual validation to ensure accurate data extraction.

[4] Another study explored an RPA-based approach for automating handwritten insurance claims digitization. Combining OCR for text recognition with RPA workflows to validate and organize data, the study found efficiency improvements but noted the importance of template-based extraction to manage form layout variations.

These findings highlight both the potential and limitations of current technologies, emphasizing the value of combining RPA and OCR to improve the efficiency and accuracy of insurance document processing.

2.3 Survey on Handwritten Text Recognition Using OCR :

The field of handwritten text recognition has evolved significantly, combining traditional OCR methods with modern AI techniques to improve accuracy and scalability. Recent research focuses on addressing challenges such as variability in handwriting styles, overlapping characters, and document quality issues. The following summarizes key studies in this domain:

[1] A study by the Journal of Imaging explored hybrid models integrating deep learning techniques with traditional OCR systems for handwritten text recognition. The researchers demonstrated improved accuracy, particularly in recognizing cursive and non-standardized handwriting, by using Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

[2] Research by the University of California highlighted the role of transfer learning in enhancing OCR performance for handwritten documents. The study leveraged pre-trained AI models to recognize text from historical and contemporary handwritten forms, achieving remarkable accuracy even with limited training data.

2.4 Summary of the intersection of RPA, AI Detection, and Plagiarism Checks:

The **Digitalization of Handwritten Insurance Forms Using UiPath** project integrates RPA, OCR, and manual validation to tackle the challenges of processing handwritten financial documents. RPA automates the workflow, ensuring consistent extraction and processing, while OCR technology identifies and digitizes handwritten data. To ensure accuracy, manual validation steps are incorporated, allowing for error correction and quality assurance.

This innovative combination addresses the challenges posed by varying handwriting styles and inconsistent document structures, ensuring a robust and efficient solution for institutions. By automating repetitive tasks and minimizing errors, the project significantly enhances operational efficiency and accuracy, positioning it as a cutting-edge solution in the domain of insurance document processing.

CHAPTER 3

SYSTEM DESIGN

3.1 SYSTEM FLOW DIAGRAM

A flowchart is a type of diagram that represents an algorithm, workflowor process. The flowchart shows the steps as boxes of various kinds,and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem. The system flow diagram for this project is in Fig. 3.1.

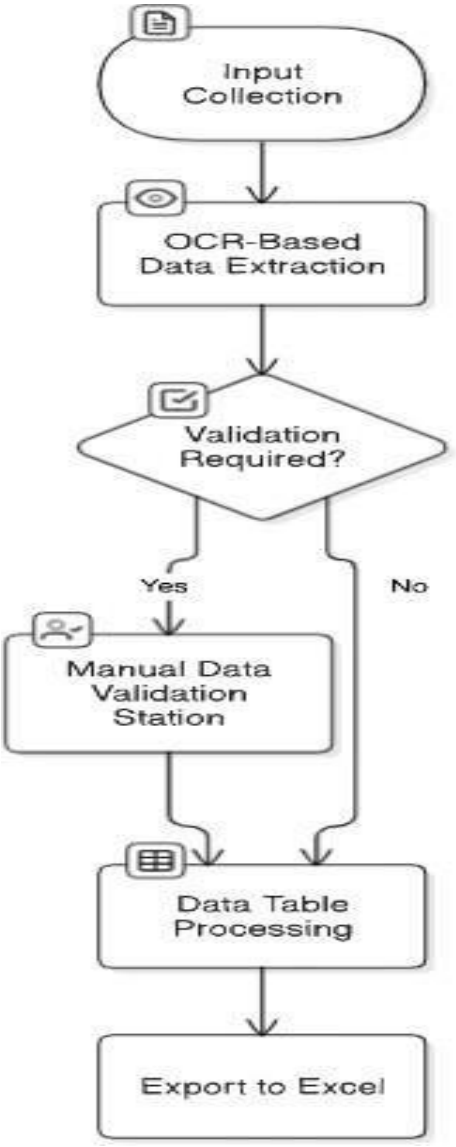


Fig 3.1 System Flow Diagram

3.2 ARCHITECTURE DIAGRAM

An architecture diagram is a graphical representation of a set of concepts, that are part of an architecture, including their principles, elements and components. The architecture diagram for this project is in Fig. 3.2.

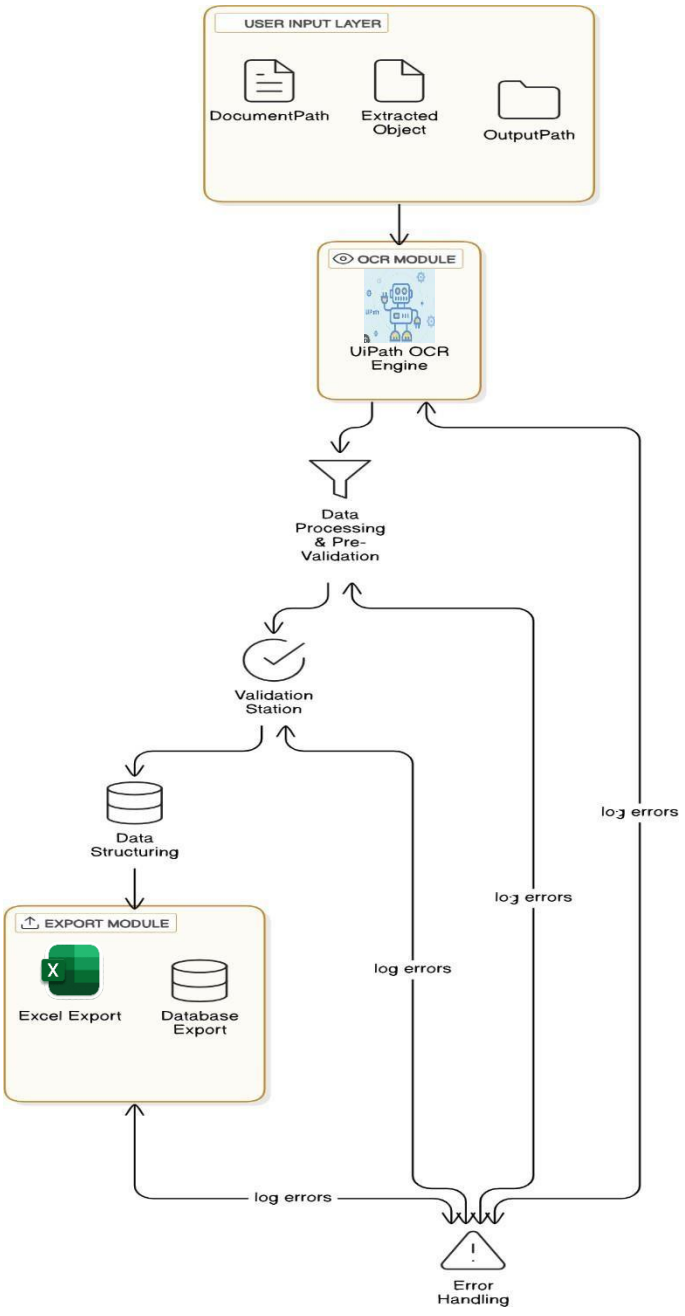


Fig 3.2 Architecture Diagram

3.3 SEQUENCE DIAGRAM

A sequence diagram is a type of interaction diagram because it describe and s how in what order a group of objects works together. Thesequence diagram for this project is in Fig. 3.3.

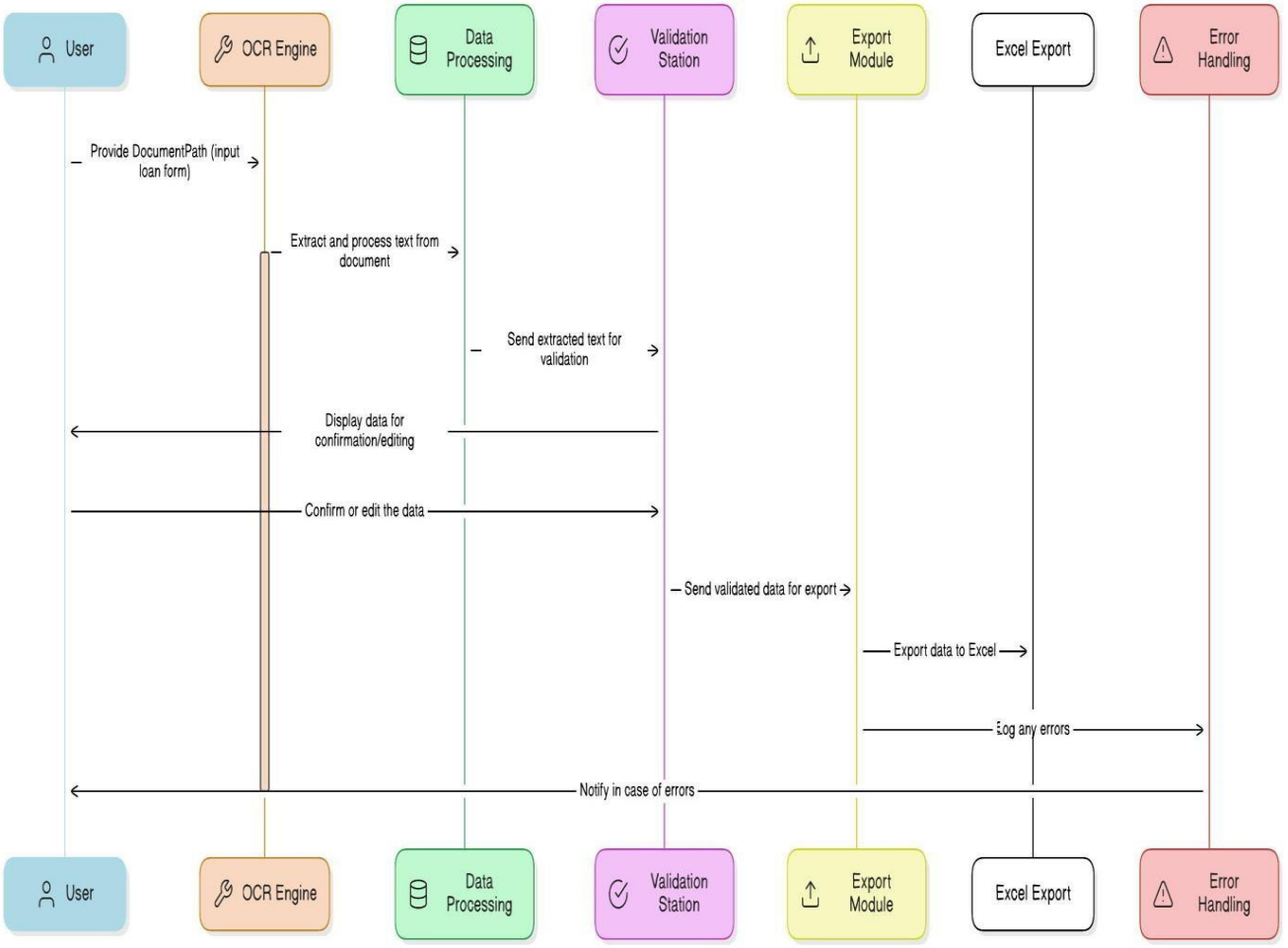


Fig 3.3 Sequence Diagram

CHAPTER 4

PROJECT DESCRIPTION

"Handwritten to Text Form Extraction in UiPath" is a Robotic Process Automation (RPA) project that automates the extraction of handwritten data from insurance documents and converts it into a digital format. Using UiPath, this system efficiently processes insurance forms, transforming handwritten details into structured, machine-readable text. This automation reduces errors, saves time, and enhances the efficiency of insurance document management, allowing for faster claims processing and policy handling while ensuring accuracy and scalability.

4.1 MODULES:

4.1.1 INPUT HANDLING AND INITIALIZATION:

4.1.1.1. Folder Selection:

- The user is prompted to select the parent folder containing insurance documents(handwritten forms) that need to be processed.
- This allows the system to target the correct folder containing the images of the loan forms for further processing.

4.1.1.2. Document Selection:

- The system automatically lists all the loan form documents within the selected folder.
- The user can choose which specific loan form(s) to process for further data extraction.

4.1.1.3. Output Path Configuration:

- The user is prompted to specify the destination (Output Path) where the processed data will be saved.
- The extracted data can be saved into structured formats like Excel sheets or a database.

4.1.2 DOCUMENT PROCESSING AND TEXT EXTRACTION:

4.1.2.1. OCR (Optical Character Recognition):

- Use UiPath's integrated OCR technology to scan and extract handwritten text from the loan documents (images).
- Various OCR engines (e.g., Google OCR, Tesseract OCR) can be used to convert the images of the handwritten forms into machine-readable text.

4.1.2.2. Data Cleansing and Validation:

- Clean and pre-process the extracted text data to correct common OCR errors (e.g., misread characters, extra spaces).
- Validate the extracted fields such as names, loan amounts, and addresses to ensure accuracy and consistency before proceeding.

4.1.3 RESULT MANAGEMENT:

4.1.3.1. Data Structuring and Storage:

- After validation, the data is organized into structured formats such as DataTables or JSON, making it easy to handle and export.
- The structured data is saved to the specified output destination (Excel file or database).

4.1.3.2. Real-time Progress Updates:

- Display real-time progress during the OCR extraction process, showing users the current status of data processing (e.g., extraction complete, validation in progress).

4.1.4 COMPLETION AND REPORTING:

4.1.4.1. Completion Message:

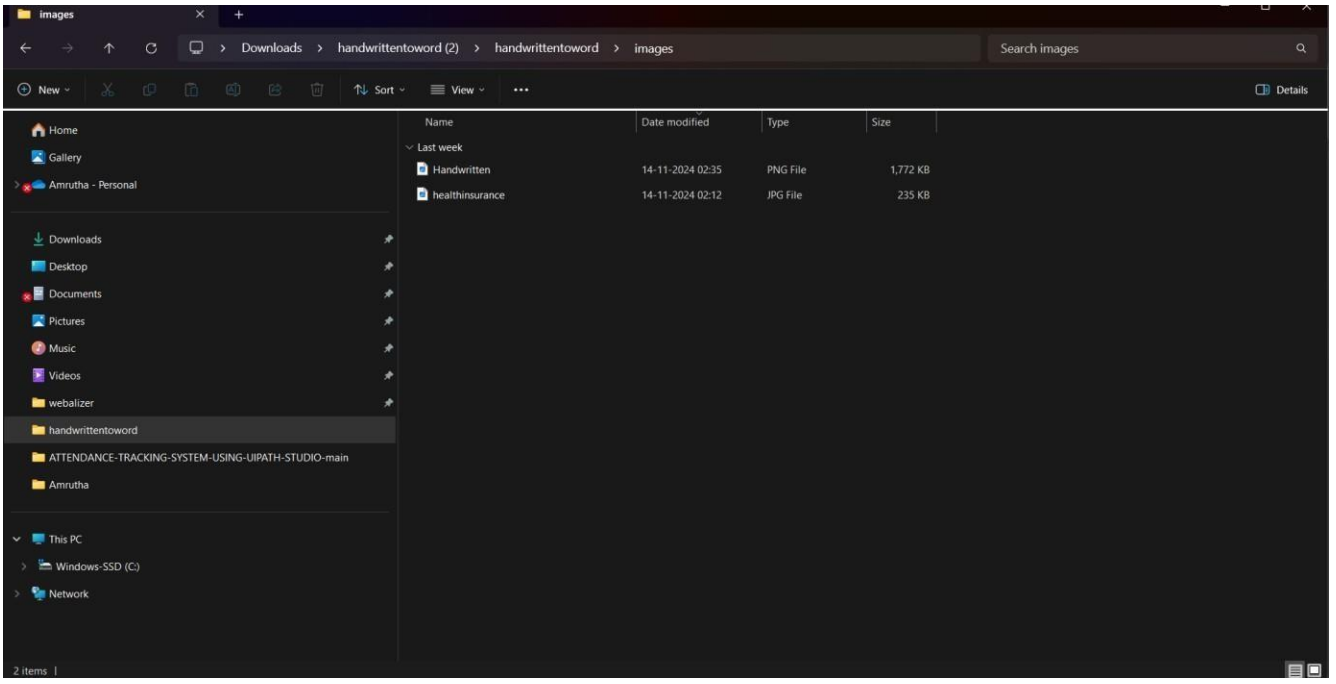
- Once the processing is complete, the system will display a message indicating the successful extraction and storage of data.
- This ensures the user is notified when all loan forms have been processed and saved.

4.1.4.2. Excel Report Generation:

- The extracted and validated data is compiled into an Excel report for easy review by the user.
- The report includes key fields such as applicant name, loan amount, address, and other relevant data, allowing users to verify and analyze the results.

CHAPTER 5

OUTPUT SCREENSHOTS



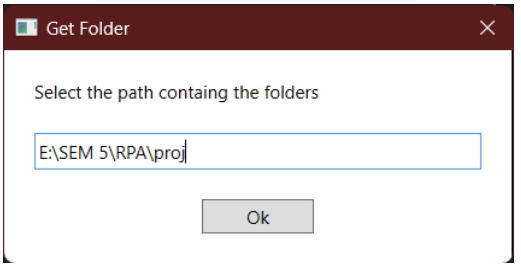


Fig 5.1 – Input Dialog

The bot retrieves the parent directory, and the user selects the folder containing the handwritten loan documents, as shown in Fig 5.1.

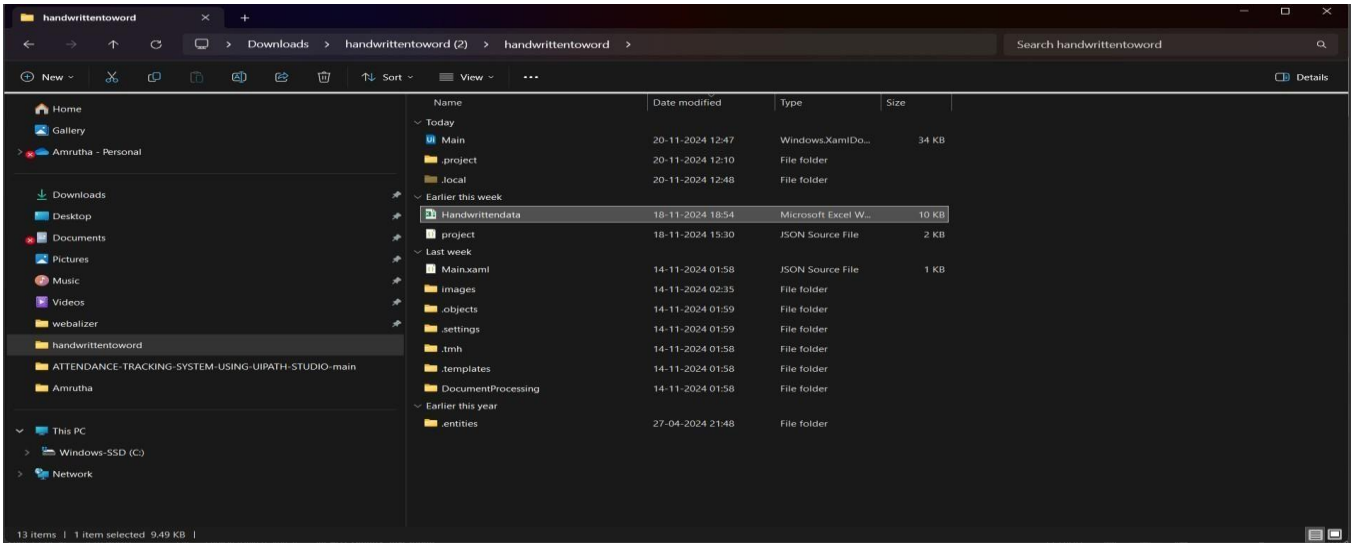


Fig 5.2 – Excel File Creation

The bot creates an excel file report in the main directory for the selected folder as shownin Fig 5.2.

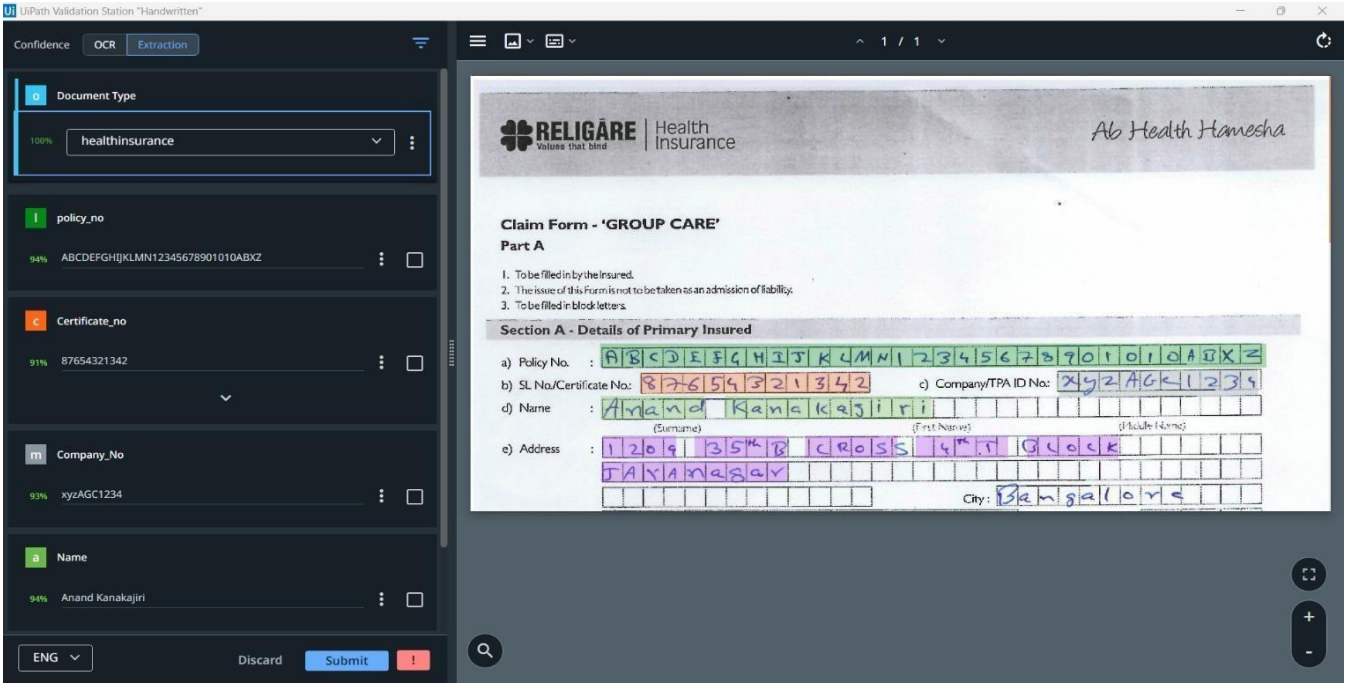
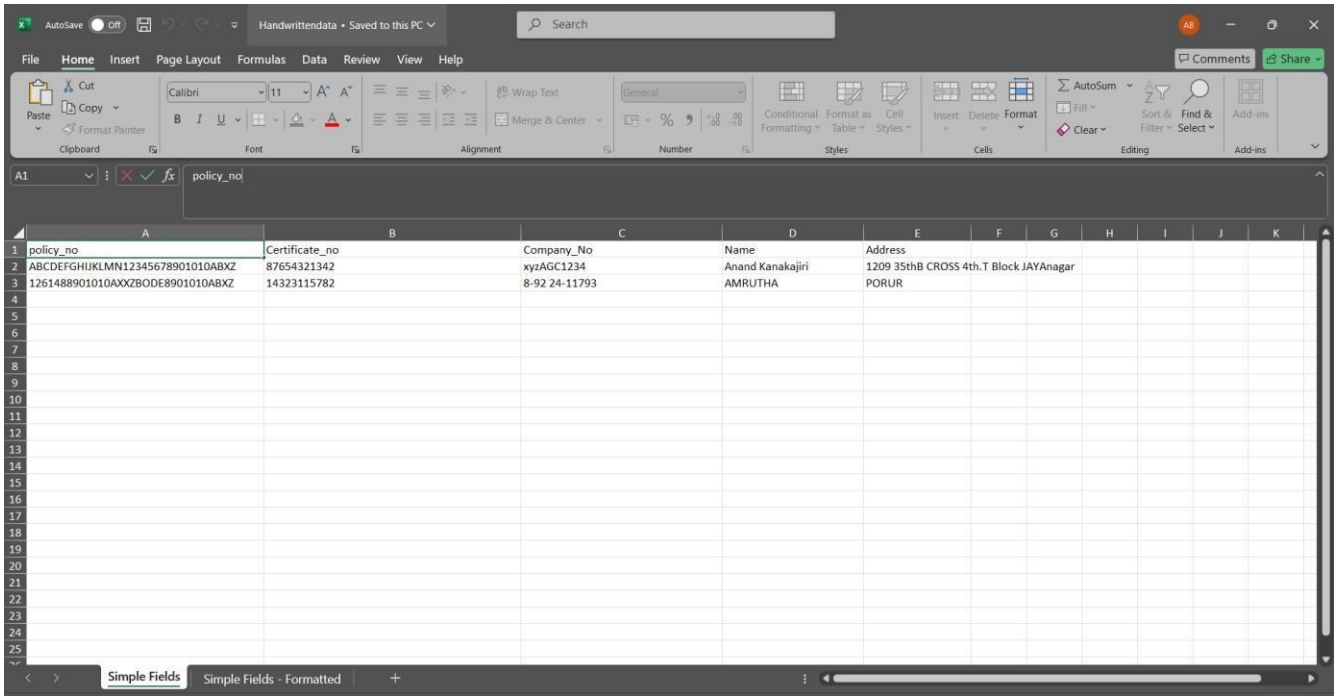


Fig 5.3 – Validation Station

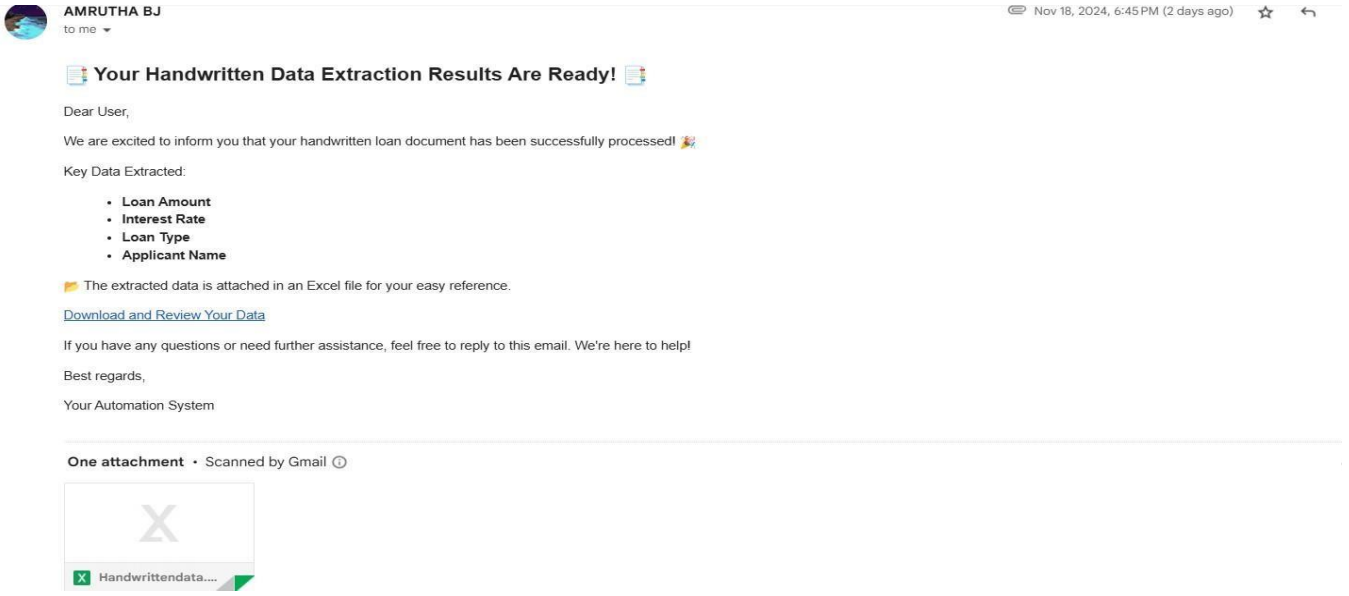
The Validation Station is an interactive interface where the bot displays the recognized text extracted from the document. Users can review the extracted data and make corrections or edits if necessary. This ensures that any errors introduced during the OCR process are resolved before proceeding to the next steps. This feature enhances accuracy and user control over the data extraction process.



policy_no	Certificate_no	Company_No	Name	Address
ABCDEFGHIJKLMN12345678901010ABXZ	87654321342	xyzAGC1234	Anand Kanakajiri	1209 35thB CROSS 4th.T Block JAYAnagar
1261488901010AXXZBODE8901010ABXZ	14323115782	8-92 24-11793	AMRUTHA	PORUR

Fig 5.4 – Excel Report

The results are then updated to the excel file that was created at the early steps of execution andsaved as it is shown in Fig 5.5.



AMRUTHA BJ
to me

Nov 18, 2024, 6:45 PM (2 days ago)

📑 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

Handwrittendata....

Fig 5.5– Email Notification

The results are sent via email, with the generated Excel file containing the extracted data attached, asshown in Fig 5.6.

CHAPTER 6

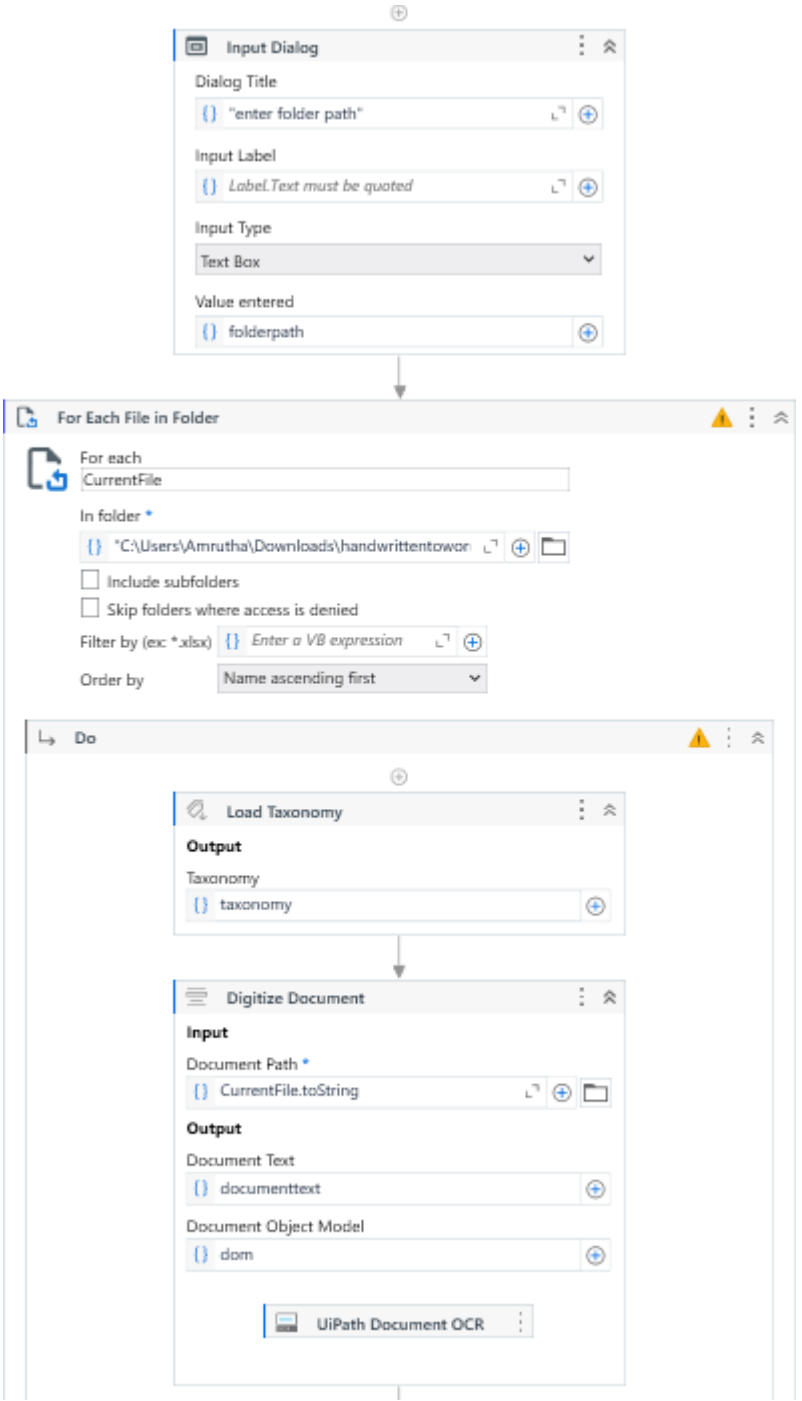
CONCLUSION

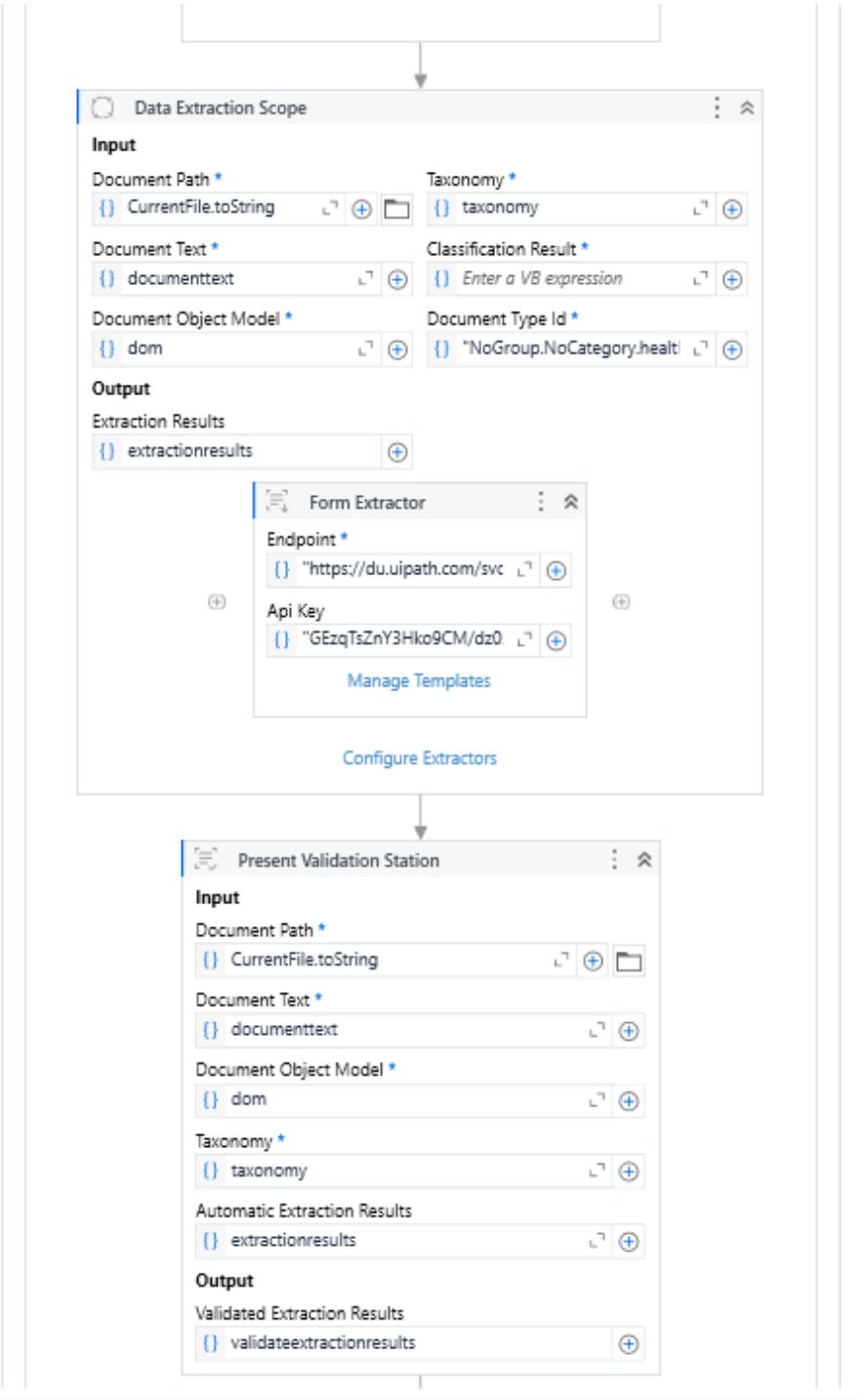
The "Smart Handwritten Insurance Document Digitization Bot" leverages UiPath's Robotic Process Automation (RPA) to automate the extraction and digitization of handwritten insurance forms. The system converts handwritten content into structured, machine-readable data, significantly reducing the time and effort required for manual data entry. This not only enhances the accuracy of data processing but also improves efficiency, ensuring faster document handling and reducing operational costs in the insurance industry.

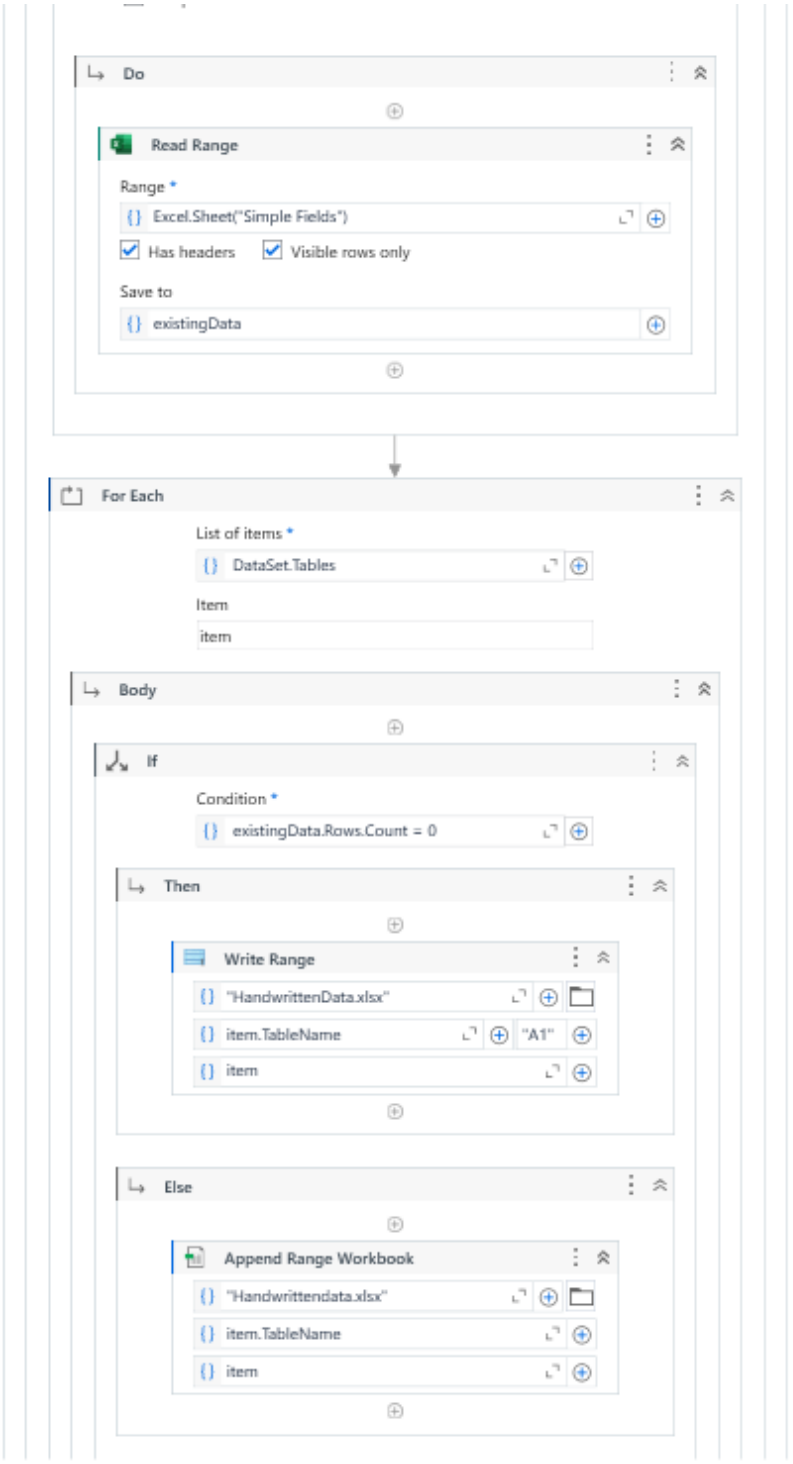
With real-time updates, the bot improves transparency and enables users to easily validate extracted data, which is recorded in a detailed Excel report. Despite challenges such as poor handwriting and context-specific issues, the system's ability to automate repetitive tasks allows insurance professionals to focus on critical decision-making. Continuous improvements in OCR algorithms and AI models ensure that the bot adapts to diverse handwriting styles, contributing to ongoing advancements in digital transformation in the insurance sector.

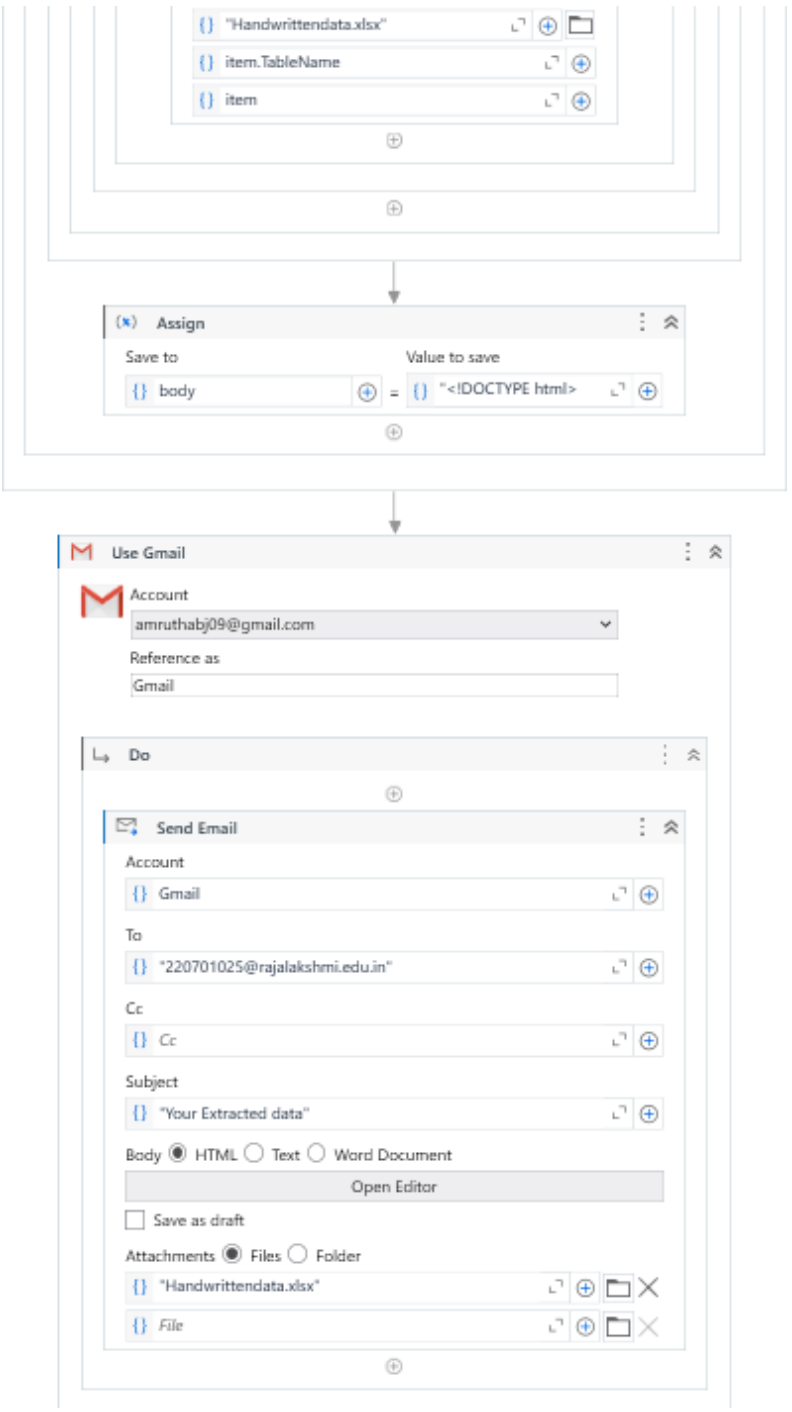
APPENDIX

PROCESS WORK FLOW




Data Extraction Scope
Input
Document Path *
CurrentFile.toString
Taxonomy *
taxonomy
Document Text *
documenttext
Classification Result *
Enter a VB expression
Document Object Model *
dom
Document Type Id *
"NoGroup.NoCategory.healt
Output
Extraction Results
extractionresults
Form Extractor
Endpoint *
"https://du.uipath.com/svc
Api Key
"GEzqTsZnY3Hko9CM/dz0
Manage Templates
Configure Extractors
Present Validation Station
Input
Document Path *
CurrentFile.toString
Document Text *
documenttext
Document Object Model *
dom
Taxonomy *
taxonomy
Automatic Extraction Results
extractionresults
Output
Validated Extraction Results
validateextractionresults


Do
Read Range
Range *
Excel.Sheet("Simple Fields")
Has headers
Visible rows only
Save to
existingData
For Each
List of items *
DataSet.Tables
Item
item
Body
If
Condition *
existingData.Rows.Count = 0
Then
Write Range
"HandwrittenData.xlsx"
item.TableName
"A1"
item
Else
Append Range Workbook
"Handwrittendata.xlsx"
item.TableName
item


"Handwrittendata.xlsx"
item.TableName
item
Assign
Save to
Value to save
body
"<!DOCTYPE html>
Use Gmail
Account
amruthabj09@gmail.com
Reference as
Gmail
Do
Send Email
Account
Gmail
To
"220701025@rajalakshmi.edu.in"
Cc
Cc
Subject
"Your Extracted data"
Body HTML Text Word Document
Open Editor
Save as draft
Attachments Files Folder
"Handwrittendata.xlsx"
File

REFERENCES

[1] "Robotic Process Automation in Insurance for Claims Processing and Fraud Detection" – This IEEE publication explores how RPA can automate repetitive tasks in the insurance industry, particularly in claims processing and fraud detection, enhancing efficiency and accuracy.

[2] "Innovating the Insurance Industry with RPA" – A white paper by Roboyo, focusing on how RPA is transforming the insurance process, especially in areas like document validation, policy management, and customer interactions, improving operational efficiency.

[3] "Enhancing Insurance Process Efficiency Through RPA" – A Forbes Insights report that discusses how RPA can streamline policy management, claims handling, and compliance, helping insurance companies offer faster, more accurate services while reducing costs.

.