ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/himadeep-movva/530

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 1

Research Article

Enhancing Robotic Process Automation with Optical Character Recognition: A Comparative Analysis of OCR Tools in UiPath

Himadeep Movva*

Citation: Movva H. Enhancing Robotic Process Automation with Optical Character Recognition: A Comparative Analysis of OCR Tools in UiPath. *J Artif Intell Mach Learn & Data Sci 2022*, 1(1), 2466-24670. DOI: doi.org/10.51219/JAIMLD/himadeep-movva/530

Received: 03 October, 2022; **Accepted:** 28 October, 2022; **Published:** 30 October, 2022

***Corresponding author:** Himadeep Movva, Independent Researcher, USA, E-mail: movvahimadeep@gmail.com

A B S T R A C T

Robotic Process Automation (RPA) has revolutionized how business processes are handled. While UI automation is mostly reliable, the automation that involves accessing documents, PDF or otherwise, has also become equally reliable, if not more, with the advent of and frequent development of image-based techniques or OCR (Optical Character techniques). This paper delves deep into the capabilities of OCR in UiPath. The potential of RPA beyond UI automation can be leveraged by realizing the true potential of OCR-based techniques. These techniques are useful for organizations that deal with semi-structured and unstructured data.

Keywords: Machine learning, RPA, OCR (Optical Character Recognition), algorithms, NLP, Document understanding, Structured data, Unstructured data, Deep learning, Native documents, Image-based documents, DOM (Document Object Model), Taxonomy, Digitization, Classification, Validation, Extraction

1. Introduction

UiPath OCR activities are pivotal in retrieving data from image-based or free-text documents. While UiPath has inbuilt OCR engines, it can also integrate with external plugins like Abby Flexicapture. Optical Character Recognition extracts printed characters from an image and transforms them into text. Based on the use case and the efficiency of the extracted results obtained after a thorough trial and error approach, the type of OCR engine needs to be selected. This paper presents the information and capabilities of different OCR engines. Also, to handle handwritten or image- based unstructured or semi-structured text, UiPath offers state of the art solution called Document understanding. In this research study, different components of document understanding are discussed and what each component does in terms of the whole document understanding process is mentioned. Furthermore, some considerations while using ocr activities are mentioned in this research study.

2. Role of OCR, Benefits and Challenges

Through OCR, conversion to electronic format is possible, letting RPA search for editable text from image-based or pdf-based documents. Automating free text or native documents might be relatively simple, but automating non-native documents or PDFs requires the usage of OCR-based techniques and/or extensive configuration. This includes the usage of ML or deep learning. Since traditional UI-based automation depends on structured, selectable text, OCR allows us to interact with unstructured and semi-structured data that might otherwise require manual intervention. One of the benefits of OCR capabilities includes eliminating manual data entry. Often, the data from the invoices or documents is extracted manually and stored in certain sheets. Then the data is provided as input to complete certain transactions within different applications in an organization. This requires much human time and may be prone to errors. The other benefit would be scalability[1].

Organizations often deal with a humongous number of documents on a day-to-day basis. Suppose the organization is within healthcare, insurance, legal services, supply chain or financial services. In that case, the number of documents and the process of mapping data from the documents with the internal system also increases. The other benefit, which is most crucial in terms of compliance and reporting, is data accuracy. When skimming through the data and entering the required data into systems is often tasked with a human, there can often be human errors, such as entering incorrect data into the internal systems. This could create incorrect reports and error-prone balance sheets. While OCR combined with RPA can handle these requests with the highest precision compared to humans, exception records are routed for human analysis. While the data, which is structured, can easily be handled and processed, processing unstructured data requires the usage of OCR along with the configuration of ML algorithms.

2.1. Below are some of the use cases of RPA and OCR integration

Invoice processing: RPA combined with OCR can extract relevant information such as vendor information and amount and send the data for further processing, including capturing information into financial systems. Traditional invoice processing often suffers from inefficiencies caused by manual data entry, which results in delays, inaccuracies and higher operational costs. By combining RPA with OCR, firms can streamline and increase invoice processing efficiency. Using Optical Character Recognition (OCR) technology, bots can retrieve information from different fields in different formats. The type of documents can be anything: pdf documents, scanned images and email attachments. After data is retrieved, if there is any discrepancy, that document can be flagged for human review. However, if all the information is correct, the Bot updates the information in the corresponding business systems as per the business logic.

Claims Processing: Across revenue cycle management (RCM), OCR is very helpful in extracting patient information related to claims and performing rote processes. These processes, especially in health care, under the RCM are very intensive in terms of number of hours it takes for the user to skim through the information and process the claims. After extraction of the results, bots validate the data against data in the databases, ensuring 100% accuracy prior to submission of a transaction. This process by the bot using OCR technology reduces human effort significantly and enables them to focus on more meaningful tasks that may require human judgment. By utilizing OCR, Bots can automatically process claims and various other processes under RCM, helping associates deliver the best and quickest patient care possible.

Legal documents processing: RPA with OCR can extract complex data from humungous legal documents and capture it into the internal systems. Processing legal documents is one of law firms' most data-heavy and labor-intensive operations. Legal agreements, contracts and regulatory papers often consist of extensive amounts of unstructured text, making manual handling slow and prone to significant errors. The combination of OCR and RPA allows the legal team to extract required content after skimming through extensive long list of documents spanning multiple pages. The collected data is mapped into the internal systems per the compliance. This kind of automation helps the legal teams focus more on important tasks, such as documentation required for adjudication, instead of on tedious tasks, such as skimming through a long list of documents to extract the required information.

Customer onboarding: OCR can also access needed documents and complete customer onboarding. It can also be used for various processes within the HR department. Customer onboarding is essential in banking, insurance and human resources sectors, where confirming identities and validating documents is vital. Historically, the onboarding process has relied on manually checking documents like passports, driver's licenses, bank statements and utility bills, which can be both lengthy and prone to mistakes. Using OCR with RPA, these processes in the HR department can be automated.

3. OCR Capabilities in UiPath

UiPath's OCR capabilities have evolved and significantly strengthened in accuracy and reliability[2]. However, Document Understanding in UiPath is an impressive feature, providing the use of OCR in combination with ML to automate and integrate various types of documents, which could vary in field names and data structure. This could be challenging to achieve with just the OCR activities. Still, the combination of OCR and ML in document understanding provides a way to automate such a process involving the abovementioned documents. Regular OCR can extract raw text from images or documents, but document understanding adds value by intelligently interpreting structured and unstructured data with varying layouts and formats[3]. Using NLP and OCR, it can identify and classify document types, significantly improving the efficiency of complex document-based processes[4].

The key OCR engines include[5]:

3.1. UiPath OCR

UiPath OCR has two types: UiPath Screen OCR and UiPath Document OCR. UiPath Screen OCR is free, has a standard package and can be used for basic screen extraction. However, UiPath Document OCR isn't free and requires a document understanding suite license. With the help of UiPath OCR, an integrated text recognition engine with several OCR options, users may extract text from scanned documents and photos in both organized and unstructured formats. UiPath OCR's automation platform incorporates both native and third-party engines.

3.1.1. UiPath screen OCR:

- Uses deep learning models to identify text from on-screen elements.
- Reliably extracts text from virtual machines, applications based out of Remote desktops and Citrix applications, where traditional UI selectors are unavailable.
- Recognizes text from dynamic UI elements and can work for multiple languages.
- Although limited in accuracy, it can extract content even in the case of low-resolution images or small fonts.

3.1.2. UiPath document OCR:

- Best suitable for extracting text from scanned pdfs and other unstructured documents.
- Utilizes UiPath's Document Understanding capabilities to classify and extract key information.

- Recognizes tabular data, headers, footers and multi-column formats, making it ideal for structured data extraction.
- Though accuracy varies, it can extract content when it is handwritten.

3.2. Tesseract OCR

It is open source, seamlessly integrated with UiPath and may be available with standard packages. However, it may struggle with efficiency when changing or creating complex layouts. It is restricted to simple text extraction.

- It is open-source, customizable and supports for over 100 languages.
- It can process structured and semi-structured text within the document.
- However, lower accuracy compared to AI-powered OCR engines.

3.3. Microsoft OCR

It is integrated with UiPath and can extract text from images or scanned documents. In some cases, it may be more reliable than Tesseract OCR.

- It is best suited for text extraction from structured files.
- It works well for printed and typed text.
- However, struggles with low-resolution and distorted text and the feature of recognizing handwriting is limited.

3.4. Google cloud vision OCR

It comes within UiPath activities and is useful for more sophisticated extraction, including handling complex documents, varying layouts and multi-languages

- Supports extracting text, both in the case of printed and handwritten text.
- Support is available and recognizes over 100 languages.

3.5. Abby flex capture

This third-party tool can seamlessly integrate with UiPath. Abby allows for high-accuracy text extraction in document extraction.

- Uses machine learning to improve accuracy and includes file-specific layouts.
- Supports retrieving both handwritten text and printed text.
- However, it is more expensive than other OCR engines and a customized initial setup is needed for different document types.

Please find the below Table 1 that provides a comparative analysis of each OCR engine. Each scenario is unique and the best OCR may be chosen only after testing with various OCR engines, logging the efficiency of the results and comparing the results.

Table 1: Different OCR engines and their features.

OCR Engine	Accuracy	Speed	Cost	Use Case Suitability
UiPath Screen OCR	Good	Fast	Free with UiPath license	Screen scraping, basic text extraction
UiPath Document OCR	High	Moderate	May require additional cost	Document processing, complex layouts
Tesseract OCR	Moderate	Moderate	Free, open source	Simple documents, single-language text
Microsoft OCR	Good	Fast	Free with UiPath license	Basic documents, multi-language support
Google Cloud OCR	High	Fast	Paid	Complex documents, handwriting recognition

Document Understanding framework combines UiPath document understanding OCR and ML algorithms, making a formidable solution in UiPath to handle varying layouts and unstructured and semi-structured documents [6]. When there is much variation in document types and the structure of documents, including semi-structured and unstructured documents and varying anchors or fields, using regular expressions or defining unique templates for each type of document is not practically viable or scalable. Document understanding is a formidable player in solving these challenges. The key components of Document Understanding are as follows:

Load Taxonomy: Taxonomy is a collection of document types. Fields will be captured from each document type. This is used to define a set of fields and document types so that documents of various types can be assessed and classified accordingly according to the defined templates. If you want to classify files into different document types, taxonomy should contain those types that need to be treated specially. Taxonomy contains a schema of data that will be used throughout the process. Below are the steps to Load Taxonomy.

3.6. Add Load Taxonomy activity to UiPath Studio

Define document type and define fields in the taxonomy manager. The output type is Document Taxonomy in the taxonomy variable. This variable will be used for further document understanding activities to guide data extraction.

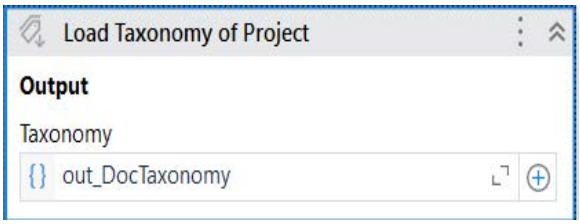

Figure 1: Load Taxonomy.

Digitization: Digitization steps convert the document into machine-readable text. This is the first step that will be applied in the Document Understanding framework. The output of this step would be a string variable that stores the text and DOM model of that file. Digitization will be useful for non-natively digital documents, including scanned PDFs, images and handwritten notes. Digitization leverages OCR to extract content from documents. The corresponding OCR engine is chosen based on the requirement. The output will be a Document Object Model (DOM). OCR does not always need to be applied, as in the case of free text documents that aren't scanned, OCR may not be required. But if the force applies OCR option is set to true, OCR will be force applied.

Document classification: This step identifies and classifies documents within the batch. This will be helpful when dealing with different types of documents, such as invoices, receipts or contracts, all at once. Once a document is classified, a specific

data extraction technique is applied to each type of document and information is retrieved effectively and with ease. There are different types of classifiers and a specific document classifier can be chosen based on the use case. The classification can be used either as an attended activity or as a type of action in the action center.

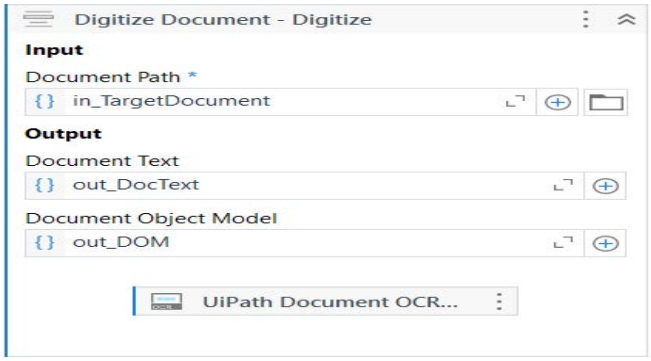

Figure 2: Digitize Document Activity.

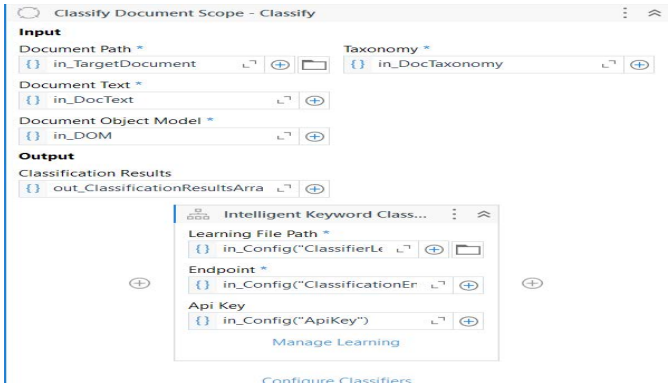

Figure 3: Classify Document Scope Activity.

Keyword-based classifier classifies documents based on keywords. For instance, invoices contain invoices as keywords in the documents. A more advanced classifier than the one mentioned above is an intelligent keyword-based classifier based on Natural Language Processing (NLP). This is more adaptable to varied document formats and languages. The third type of classifier is the Machine Learning classifier, which has a training machine learning model for each type to improve accuracy. This would be very helpful in the case of document processing, where formats and content vary widely.

c. Document Classification Validation: This step provides an option for humans to validate document classifications. This is especially crucial for certain documents with very low confidence classification. By looping human validation, accuracy is ensured to improve. Document classification validation step is very crucial as incorrect classification, which means too many pages or too few pages are classified under a classification type, leads to incorrect extraction results. Hence, 100% accuracy needs to be attained in document classification.

d. Data extraction: Place the data extraction scope activity after Classify Document Scope in the workflow

Set the Taxonomy property to reference the taxonomy variable loaded with Load Taxonomy and Set the Document Text property to the output from Digitize Document or another source of document text. Use the Document Type ID property to specify the document type (from the classification output) so the Data Extraction Scope knows which fields to extract. Extraction is done using Data Extraction Scope activity. There can be multiple extractors.

3.7. Configure extractors

Various extractors can be added inside the Data extraction scope, namely reg-ex-based extractors, ML extractors, form extractors and intelligent form extractors. Setup the field mapping and output of data extraction scope is out Extraction Results object.

Data Export: Finally, the document is exported for further use. Data extraction component extracts very specific content or fields from the documents. However, only the fields that were previously defined in the taxonomy manager can be extracted out using data extraction activity. Any field that wasn't defined in the taxonomy manger while creating fields can't be retrieved at this stage. This step should be run certain number of times based on the number of classifications. For instance, if the classification result has only one document type, then it should be run once. However, if there are multiple types of documents within the same document, the classification needs to be run those many numbers of times. Data is retrieved using data extraction scope activity. In addition, there are optional stages, which are described below.

3.8. Document extraction validation

After document extraction, one optional but recommended step is document extraction validation. This step involves a human review of extracted results and correction if required. This step ensures that the extracted results is 100% correct. Not always, this step may be configured. Sometimes, based on the availability of users to validate this step can be configured. In some cases, data can be validated against available data sources, including databases for validation. In some other cases, this may not be possible and hence require human validation of information. This step can also be configured as an attended activity or an action center task. We can use present validation station to configure this[2] steps as an attended activity. We can also create an orchestration process and assign tasks in action center by using persistence activities: Document Validation Action and pause for the Document Validation Action to complete before resuming activities, as this process links the unattended robots handling the automated tasks with the Knowledge Workers who can retrieve actions from the Action Center.

3.9. Document classification training

This document understanding (DU) framework component is used to close the feedback loop for classifiers after learning from human feedback. The training in the framework may not be required in some cases where classifiers may not support retraining when the process needs to always use the same training or when classifiers are updated outside of the DU framework. They become self-learning algorithms based on the feedback from the users. Classification training is done by using train classifiers cope activity. We can even train more algorithms in one go. This part of the framework needs to run only after document classification validation is completed, as only feedback that is confirmed by the user needs to be trained for the algorithm. Training has to be run both in the case of failed transaction where a human gives input and algorithm learns from it and in the case of successful one where the algorithm learns about the successful transaction and its data. This activity provides extractors necessary algorithms to run, allows minimum confidence threshold at field level extraction and lets us customize the number of fields from extractor. Extractors can

be configured in a hybrid approach, where certain fields can be extracted by one extractor and other fields can be extracted by another extractor. If multiple extractors are configured, they're extracted from right to left and if an extractor doesn't retrieve fields with a reasonable threshold, those fields can be extracted by another extractor.

4. Challenges and considerations

Here would be the main considerations when implementing OCR in RPA tools such as UiPath,

4.1. Accuracy and reliability

If the quality of original document is low that could have a significant impact on the performance of OCR engine. But, in case of high accuracy in documents, the OCR engines would fare best. In entirety, the quality or resolution of documents has repercussions on performance on OCR engine. The performance of OCR is often shrieked by documents such as handwritten material and low-resolution scans. It can be difficult to ensure high accuracy when dealing with papers with numerous languages, different layouts or non-standard fonts. Businesses need to weigh the trade-off between the quality of the source material they process and the dependability of OCR.

4.2. Processing unstructured data

While the OCR capabilities of UiPath, including document understanding and those of Power Automate, including AI builders, are complex to build, integration between workflows and OCR systems can be even more complex; effective integration requires seamless communication between the OCR engine and RPA processes. More often, RPA processes are tasked with retrieving information from unstructured documents and configuring workflows toward the effect of recognizing this data can be complex. It requires sophisticated document extraction techniques, which, sometimes, may not be fully supported by the RPA tool.

4.3. Resource intensity and costs

OCR implementation is quite resource intensive, both in terms of time to develop and test continuously and may require significant processing power for a large volume of documents, further escalating development costs.

4.4. Data security and privacy

OCR processing involves extracting sensitive information from documents across various industries, including healthcare, finance and legal. It is essential that the data is securely handled and compliance standards, including GDPR and HIPAA, are met. Equally important and in some cases, more so, is assessing whether OCR processing should be conducted in an on-premises secured cloud-based environment or a hybrid approach.

4.5. Error handling and exception management

OCR engines are not 100% accurate and there could be slight degree of error while extracting information from documents. To handle this shortcoming, an attended automation framework including a human-in-the-loop solution is essential so that errors can be assigned to users for fixing. This adds further complexity to the already challenging factors in developing OCR workflows.

4.6. Impact on RPA workflow performance

Using desktop automation, UI automation or any traditional automation with robust elements to identify and configure elements reliably, efficiency would be higher, sometimes more than 90% of successful transactions. However, with the introduction of OCR into the RPA workflows, a further decrease in efficiency, sometimes a significant decrease when compared with general cases, may be expected and careful analysis must be made with respect to ROI.

4.7. Licensing and cost structure of OCR tools

Many OCR engines may charge based on the volume of transactions, significantly increasing the costs of processing transactions in case of large volumes of transactions. Budgeting ongoing OCR costs, especially when high throughput is required, is essential. Organizations need to consider cost structure carefully when selecting OCR tools.

4.8. Evaluation and testing for accuracy

Not all OCR engines work efficiently for all cases of automation. Each OCR engine or technique may be assessed and tested with output results and needs to be tried on a trial and error-based approach to determine what OCR technique works better for the current use case at hand.

5. Conclusion

In conclusion, we can say that prior to the availability of document understanding and robust OCR engines, there was limitation in extracting content from nonnative documents or scanned images. However, a combination of OCR and Document understanding has created a robust attended automation or orchestration framework where information from even semi-structured and unstructured documents can be retrieved with reasonable accuracy. In some cases, the extraction results from Microsoft OCR engine are more dependable than those produced by Tesseract OCR. However, there are also situations where Tesseract OCR yields superior results. Hence, the usage of different OCR algorithms depends on the corresponding use case[7]. However, deploying Optical Character Recognition (OCR) is not straightforward and is fraught with many challenges, including processing costs, licensing, throughput and data privacy. The organization and a team of architects need to carefully assess the throughput and development efforts before strategizing to continue with the Optical Character Recognition (OCR) based approach.

6. References

1. Cutting GA, Cutting-Decelle A, Intelligent Document Processing - Methods and Tools in the real world, 2021.
2. https://smartbridge.medium.com/comparing-ocr-engines-in-uipath-196333109670
3. https://docs.uipath.com/document-understanding/standalone/2021.10/user-guide/ocr-services1
4. https://www.uipath.com/product/document-understanding
5. https://docs.uipath.com/document-understanding/standalone/2022.4/user-guide/du-ocr-services
6. https://forum.uipath.com/t/document-understanding-best-practises-reframwork/260892
7. Malathi T. An Experimental Performance Analysis on Robotics Process Automation (RPA) With Open Source OCR Engines: Microsoft, rspsciencehub, 2021;1059.