*Telehealth Ecosystems in Practice*
*M. Giacomini et al. (Eds.)*

*doi:10.3233/SHTI230741*

# Robotic Process Automation Based Data Extraction from Handwritten Medical Forms

Norbert GAL-NADASAN [a,1] Vasile STOICU-TIVADAR[a], Emanuela GAL-NADASAN[b] and Anca Raluca DINU[b]
[a] *Politehnica University of Timisoara, Department of Automation and Applied Informatics, Faculty of Automation and Computers, Romania*
[b] *Department of Medical Rehabilitation, University of Medicine and Pharmacy "Victor Babes"*
*ORCiD ID: Gal-Nadasan Norbert https://orcid.org/0000-0002-5690-2674*

**Abstract.** This paper proposes to create an RPA(robotic process automation) based software robot that can digitalize and extract data from handwritten medical forms. The RPA robot uses a taxonomy that is specific for the medical form and associates the extracted data with the taxonomy. This is accomplished using UiPath studio to create the robot, Google Cloud Vision OCR(optical character recognition) to create the DOM (digital object model) file and UiPath machine learning (ML) API to extract the data from the medical form. Due to the fact that the medical form is in a non-standard format a data extraction template had to be applied. After the extraction process the data can be saved into databases or into a spreadsheets.

**Keywords.** Robotic Process Automation (RPA), Optical character recognition (OCR), handwriting, data extraction

## 1. Introduction

Robotic Software Automation (RPA) is a technology that creates software robots which can perform repetitive human tasks [1,2]. RPA robots are constructed using predefined activities that mimic human computer interaction by reducing the manual repetitive work and errors to a large extent [3]. The most common RPA development tools are UiPath, Automation Anywhere and Blue Prism.

The most use cases of using Robotic Process Automation can be found in the domains where many files must be processed like secretary productivity [4], document processing automations [5,6] and document classification [7].

In one instance in the medical domain RPA and machine learning (ML) is used to detect the early signs of Glaucoma [8]. In another instance RPA is implemented in elderly friendly consumer services [9] and chat bots [10]

Our idea is to create an RPA based software robot that helps physicians and other healthcare professionals to digitize handwritten medical forms. Even though the year is

[1] Norbert GAL-NADASAN, Politehnica University of Timisoara, Department of Automation and Applied Informatics, Faculty of Automation and Computers, E-mail: norbert.gal@upt.ro.

2023 many small medical healthcare facilities utilize handwritten forms and many older medical forms and other forms are in handwritten formats. This solution will provide a base for further development of robots that can bulk process a large number of forms for digital archiving purposes.

## 2. Using RPA and OCR to process handwritten patient forms.

In our country small local medical clinics and cabinets still use handwritten medical forms. Generally, these forms have a predefined structure, and the data is filled in by hand. These forms are passed on from the medical doctor to the medical assistant or rehabilitation practitioner by hand and are archived in physical cabinets.

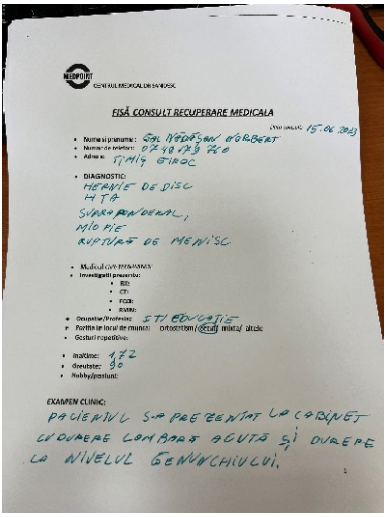


**Figure 1.** Handwritten medical form (first author's fictive test data).

The problem arouses when these handwritten medical forms must be archived in digital form. For this a reliable batch processing solution must be found that can digitize the physical documents with minimal human intervention.

Our solution proposes the usage of RPA by creating a software robot that can digitize the documents and extract useful information in a structured manner.

To develop the RPA robot UiPath studio was used with community type licensing.

### *2.1. Building the RPA Robot for data extraction*

The intended usage of the RPA robot is to extract and structure data from handwritten medical forms. These forms have, as can be observed, a non-standard format and so it can be difficult to digitize and extract meaningful data.

The first step in building the RPA robot is to create a taxonomy of the form. Taxonomy helps the robot to correlate the extracted. The taxonomy was created using UiPath's Taxonomy Manager. Figure 1 displays the created taxonomy for the medical form. This is loaded into the RPA robot using the Load Taxonomy activity.

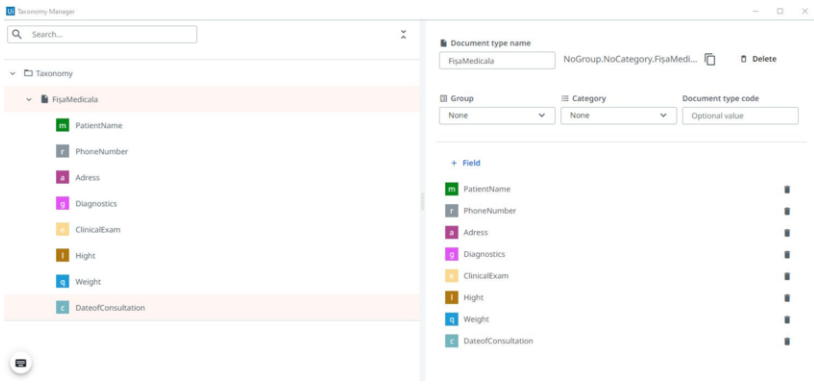


**Figure 2.** The taxonomy of the medical form.

The next phase is to create a digital version of the document that will be processed. This document can be in PDF form or an in a standard bitmap image form (jpg, jpeg, npg, bmp, ect.). This is done using the digitize document activity. This activity uses an OCR engine the creates a DOM type digital file that can be used for data extraction. The chosen OCR was the Google Cloud Vison OCR. In our tests to determine the best OCR this performed the best to recognize handwritten texts. If needed google cloud has a dedicated handwriting recognition OCR.

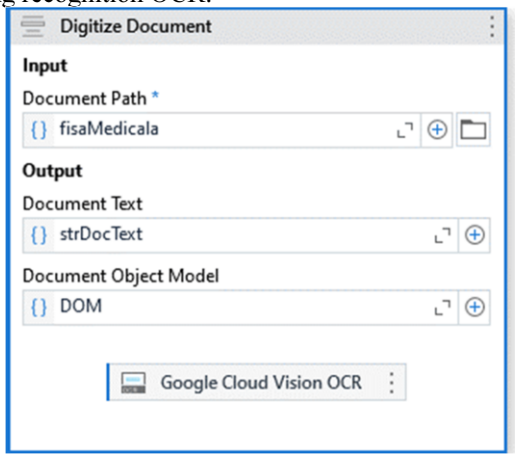


**Figure 3.** The document digitalization activity.

After the digitalized document is obtained, we can proceed to classify the document as a medical form document with the Classify Document scope activity where a keyword based classier is invoked. Basically, this classifier searches within the document for predefined keywords to ensure that the data extraction is accomplished on the right document. If the classification is accomplished correctly the Data Extraction Scope is invoked. In this scope the Form Extractor activity is used which uses a ML algorithm (provided by UiPath) to identify the date associated with the taxonomy. This for extractor activity needs a predetermined form template to identify the location of the data associated to the taxonomy. The predetermined form template is mandatory since handwritten medical form is a nonstandard form. As a drawback if the structure of the analyzed form changes the form template must be changed as well.

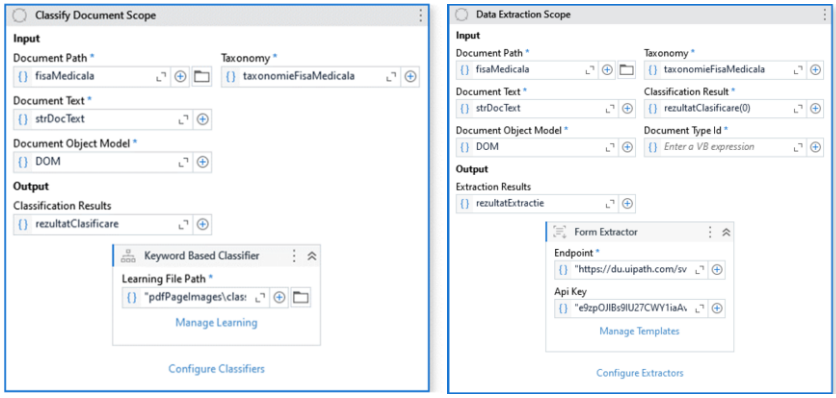


**Figure 4.** Document classification and data extraction.

## 3. Experimental Results

To view the extraction results a Presentation Validation Station activity was used presented in figure 6. This activity generates a graphical interface that presents the extracted data and its location on the form, as well as the extraction confidence values. It can be observed that all the targeted data was extracted correctly.

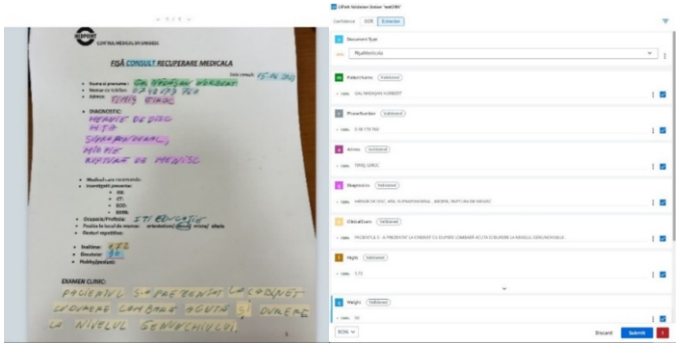

**Figure 5.** The extracted data and extraction confidence values.

The extracted data is saved using the Export Extraction Results activity into a DataSet.

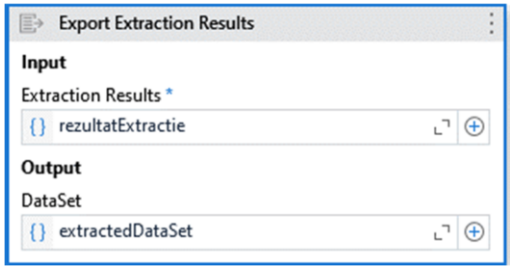


**Figure 6.** Export Extraction Results activity.

The exported data set can be used to write the extracted data into a spreadsheet type file (Excel file, Google spreadsheet, etc…), generate text type pdf's, to save the data into a database.

## 4. Discussion and Conclusions

On major factor that affects the data extraction quality is the actual handwriting of the person who wrote the form. If the handwriting is in capitalized letters the extraction confidence is 100% as shown in figure 5, but if the handwriting is a cursive un-organized handwriting the confidence can drop under 50% and then a new AI must be trained for each medical staff's handwriting.

Another aspect that needs to be addressed is the number of pages of the form. If the form have several pages a page template needs to be created for each page apart. Basically in the document classification part each page must be classified individually and the correct template to be applied. By creating several templates the proposed method can be applied to digitize any kind of hand written forms, medical and non-medical, speeding up the archiving process of hand written documents.

As a further development several bulk loading methods can be applied like accessing multiple files form a folder, from a cloud drive ex: Google drive.

## References

[1] Siderska J. Robotic Process Automation—a driver of digital transformation?. Engineering Management in Production and Services. 2020 Jun;12(2):21-31.

[2] Dube T, Eck RV, Zuva T. Review of technology adoption models and theories to measure readiness and acceptable use of technology in a business organization. Journal of Information Technology and Digital World. 2020 Dec 5;2(4):207-12.

[3] Tracy RP. IT security management and business process automation: Challenges, approaches, and rewards. Information Systems Security. 2007 Jan 1;16(2):114-22.

[4] William W, William L. Improving corporate secretary productivity using robotic process automation. In2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI) 2019 Nov 21 (pp. 1-5). IEEE. doi: 10.1109/TAAI48200.2019.8959872.

[5] Li J, Lin CM, Hu SX. Intelligent document processing method based on robot process automation. In2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing) 2021 Oct 15 (pp. 1-6). IEEE. 1-6, doi: 10.1109/PHM-Nanjing52125.2021.9613052

[6] Roopesh N, Babu CN. Robotic Process Automation for Resume Processing System. In2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) 2021 Aug 27 (pp. 180-184). IEEE. doi: 10.1109/RTEICT52294.2021.9573595.

[7] Baidya A. Document analysis and classification: A robotic process automation (RPA) and machine learning approach. In2021 4th International Conference on Information and Computer Technologies (ICICT) 2021 Mar 11 (pp. 33-37). IEEE. doi: 10.1109/ICICT52872.2021.00013.

[8] Thainimit S, Chaipayom P, Sa-arnwong N, Gansawat D, Petchyim S, Pongrujikorn S. Robotic process automation support in telemedicine: Glaucoma screening usage case. Informatics in Medicine Unlocked. 2022 Jan 1;31:101001. https://doi.org/10.1016/j.imu.2022.101001.

[9] Kobayashi T, Arai K, Imai T, Tanimoto S, Sato H, Kanai A. Communication robot for elderly based on robotic process automation. In2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC) 2019 Jul 15 (Vol. 2, pp. 251-256). IEEE. doi: 10.1109/COMPSAC.2019.10215.

[10] Bindiya AJ, Anitha N, Indumathi S. Robotic Process Automation (RPA): A software bot for healthcare sector. In2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE) 2023 Jan 27 (pp. 685-689). IEEE. doi: 10.1109/IITCEE57236.2023.10090996.