



HAL
open science

Model-Based Artificial Intelligence Architecture for Digitizing Handwritten Medication Error Reports

Mohamed Ayachi Brini, Hanae Touati, Rafika Thabet, Franck Fontanili, Marie-Hélène Cleostrate, Marie-Noëlle Cufi, Marc Pruski, Elyes Lamine

► **To cite this version:**

Mohamed Ayachi Brini, Hanae Touati, Rafika Thabet, Franck Fontanili, Marie-Hélène Cleostrate, et al.. Model-Based Artificial Intelligence Architecture for Digitizing Handwritten Medication Error Reports. 2024 IEEE/ACS - 21st International Conference on Computer Systems and Applications (AICCSA), Oct 2024, Sousse, Tunisia. pp.1-6, <10.1109/AICCSA63423.2024.10912536>. <hal-05002659>

HAL Id: hal-05002659

<https://imt-mines-albi.hal.science/hal-05002659v1>

Submitted on 10 Apr 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Model-Based Artificial Intelligence Architecture for Digitizing Handwritten Medication Error Reports

Mohamed Ayachi Brini
Toulouse University, CGI, IMT Mines Albi
Sousse University, ISITCom,
Albi, France
brini.mohamedayechi@gmail.com

Hanae Touati
Toulouse University, CGI, IMT Mines Albi
Albi, France
hanae.touati@mines-albi.fr

Rafika Thabet
Toulouse University, CGI, IMT Mines Albi
Sousse University, ISITCom,
MARS Laboratory, LR17RS05, Tunisia
Albi, France
rafika.thabet@mines-albi.fr

Franck Fontanili
Toulouse University, CGI,
IMT Mines Albi
Albi, France
franck.fontanili@mines-albi.fr

Marie-Hélène Cleostrate
Centre Hospitalier Intercommunal
Castres-Mazamet, France
marie-helene.cleostrate@chic-cm.fr

Marie-Noëlle Cufi
Centre Hospitalier Intercommunal
Castres-Mazamet, France
marie-noelle.cufi@chic-cm.fr

Marc Pruski
Atout Majeur Concept
Toulouse, France
mpruski@synora.fr

Elyes Lamine
Toulouse University, ISIS Engineering, INUC
Toulouse University, CGI, IMT Mines Albi
Castres, France
elyes.lamine@univ-jfc.fr

Abstract—Optical Character Recognition (OCR) is extremely useful in various sectors for exploring massive archived data. This technology enables the digitization of printed and handwritten texts that are frequently present in the medical field. For instance, medication error (ME) reports were previously and still in some healthcare facilities written manually, this has led to the accumulation of numerous handwritten data that are unfortunately challenging to exploit. Their digitization through OCR allows extracting important data from these documents and using them to populate the database to implement future analysis techniques to optimize the medication error management process. This paper presents a transformer-based handwritten recognition architecture that employs the Transformer-Based Optical Character Recognition (TrOCR) model combined with image segmentation techniques. Although the TrOCR model provided by Microsoft performs reasonably well in handwritten recognition, it is limited to English text because its pre-trained version was trained exclusively on English samples. This limitation is problematic for us, as our task involves digitizing French medication dictation errors. Additionally, its limitation to processing single-line text images impairs its ability to recognize paragraphs. To address these limitations, we will fine-tune the model on French handwritten data and integrate a single-line level segmentation technique, thereby overcoming these constraints. Therefore, the preliminary results from implementing our proposed architecture are promising for the digitization of medication error reports.

Index Terms—Optical Character Recognition (OCR), Deep Learning (DL), Transformer, Medication Error Management

I. INTRODUCTION

Patient safety is an absolute priority. In healthcare facilities, risk management and incident prevention are essential

for continuously improving care quality. Among the many concerns, the management of medication errors emerges as a crucial area that requires immediate attention [1]. Medication Errors (ME) are defined as any unintentional mistake during the care process involving a health product or medication, during dispensing, administration, or prescription [2]. They can have serious consequences on patient health, justifying the necessity of their prevention to reduce incident risks and improve the effectiveness of medical treatments [1].

When it comes to managing MEs, healthcare professionals rely on reports made either on paper forms or dedicated digital forms to analyze their causes and effects and propose corrective and/or preventive action plans [3] [2]. However, despite the emergence of digital tools in the medical sector, many handwritten ME reports are still archived and remain unexploited. These documents contain valuable information on past ME. If properly digitized and analyzed, they could complement the data entered in digital forms, contributing to the optimization of the ME management process and the improvement of care quality and safety. Therefore, Optical Character Recognition (OCR) technology proves to be a powerful solution for transforming handwritten documents into digital ones. [4] However, applying OCR to handwritten data presents several technical and scientific challenges, particularly concerning the processing of different handwriting styles and distinguishing between handwriting and various marks or patterns that may appear on paper, which can influence the accuracy and performance of handwritten text digitization [4]. Over the past years, the evolution of artificial intelligence (AI)

has become increasingly important across various sectors and domains such as computer vision, Natural Language Processing (NLP), Time series forecasting, etc. [5] When it comes to OCR, artificial intelligence (AI) has demonstrated impressive performance in recognizing both printed and handwritten text across a wide range of applications. We observe a variety of model architectures designed for OCR, including deep learning-based architectures that combine Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [6] [7] [8], as well as Transformer-based architectures [9] [10] [11], which has demonstrated state-of-the-art performance in handwritten Recognition. In this proposed study, we aim to evaluate the performance of artificial intelligence, specifically Transformer architectures, in digitizing medication error reports. Our approach will be based on the Transformer-based Optical Character Recognition (TrOCR) model [9] which is a transformer dedicated for handwritten text recognition. This model reached state of the art performance in printed and handwritten text recognition. Despite its notable performance, the pre-trained version TrOCR model provided by Microsoft has limitations, including support for English-only texts and the constraint of processing single-line text images which is challenging in our case since we are aiming to digitize handwritten French medication error reports. Thus in this paper, we propose the fine-tuning of Microsoft’s pre-trained transformer model TrOCR, which is more efficient in terms of time, computational, and data resources compared to training a model from scratch.

This paper is structured as follows: Section 2 presents an overview of different OCR techniques. Our proposed architecture is introduced in Section 3. The preliminary results of the digitization of MEs are presented in Section 4. Finally, Section 5 presents our conclusion and future works.

II. RELATED WORK

A. Optical character recognition architectures

Optical Character Recognition (OCR) is a process that allows reading printed text and converting it into a format that a computer, such as ASCII, codes can process. With an OCR system, users can scan magazine articles, receipts, or printed tickets, and convert them into files that can be manipulated by word processing software, for example. Advanced OCR systems can read texts in different digital fonts [12]. When it comes to applying OCR to handwritten documents, it involves converting documents written manually by different authors into editable and readable digital text.

Despite the performance of modern techniques, extracting handwritten text presents a difficult challenge. Applying OCR techniques on handwritten documents to extract the text is a major obstacle in the field of artificial intelligence (AI): stylistic variation, image quality, and the condition of the paper are factors that increase the complexity of processing archived handwritten data. Fortunately, many AI models have demonstrated excellent results in various studies over recent years. Table I below presents some models and their perfor-

mance tested on the IAM database [13], which contains 13,353 images of handwritten text lines created by 657 authors.

For each architecture, the authors chose to apply an evaluation by character error rate (CER) as the main comparison metric. The TrOCR, based on the classic Transformer architecture (encoder + decoder), has shown impressive performance with a CER of only 2.89% [9], this model was later improved into a new model, DTrOCR, by [10]. This model features an architecture that diverges from the traditional Transformer, relying only on a decoder and a language model (LM) to generate text achieving a CER of 2.38%, making it the best-performing model in the table. Furthermore [6], the VAN (Vertical Attention Network) model has the advantage of being able to extract text from an entire paragraph, unlike other models limited to a single line of text. However, it has the disadvantage of presenting the highest CER at 4.45%, with a word error rate (WER) of 14.55%, which is a limitation in terms of performance. Finally, the basic CRNN architecture [8], which integrates CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network) layers, shows a CER of 9.18%, higher than the other models.

This state-of-the-art and published benchmark of OCR models prompts us to opt for the TrOCR model to build our architecture. Indeed, it exhibits high performance with a low character error rate and is also the only model that provides an open-source pre-trained version among those presented, allowing us to save resources and time.

TABLE I
EVALUATION OF DIFFERENT MODELS AND ARCHITECTURES USING THE CHARACTER ERROR RATE (CER) METRIC ON THE IAM DATASET

Model		TrOCR	DTrOCR	VAN	CRNN
Metric	CER	2,89%	2,38%	4,45%	4,57%
	WER			14,55%	

B. Transformer Based Optical Character Recognition – TrOCR

The TrOCR model [9] is based on the Transformer architecture, as shown in Figure 1. It consists of an image Transformer responsible for visual feature extraction and a text Transformer responsible for linguistic modeling. The structure adopted by TrOCR is the classic Transformer model’s encoder-decoder structure. The encoder is designed to capture the representation of image segments, while the decoder aims to generate the sequence of word pieces, guided by visual features and previous predictions. The encoder and decoder are equipped with an attention mechanism [14] that calculates a score for each word, determining its position in the generated sentence to improve the output text’s quality and make it more understandable.

The TrOCR is a pre-trained model, which is advantageous in terms of reducing the time and resources required. It offers improved prediction accuracy thanks to the attention mechanism integrated into its classic Transformer architecture. However, since it was trained in English handwritten texts, its

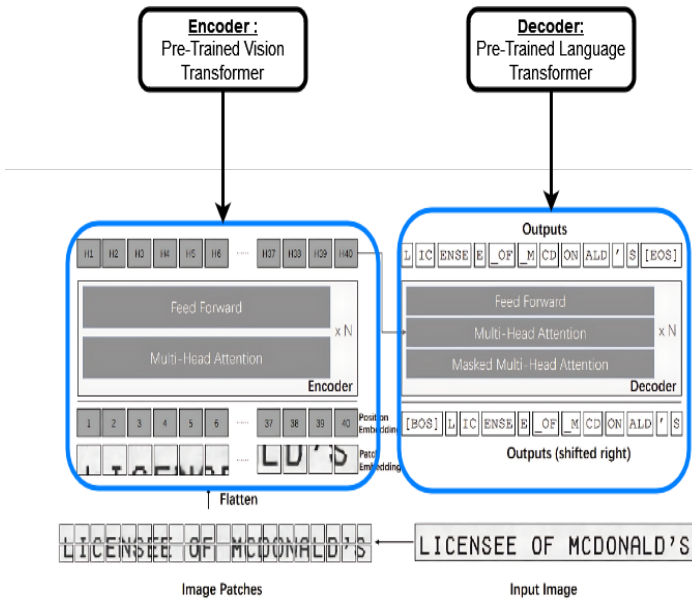


Fig. 1. TrOCR Model Architecture [9]

performance is limited when it comes to French handwritten texts. Furthermore, the efficiency of this model requires that each submitted input must be a single line of text image, thus requiring initial segmentation of the image into multiple sub-images. Each of these sub-images must include a single line of text to ensure optimized processing. In many previous studies [6], [7], instead of segmenting the image into sub-images to isolate each line of text in a paragraph, the authors used Deep Learning architectures to identify lines within paragraphs. A relevant example is the VAN model [6], which employs a Fully Convolutional Network (FCN) block to perform this task. However, although this approach is effective, it requires significant time and resources to train the model architecture.

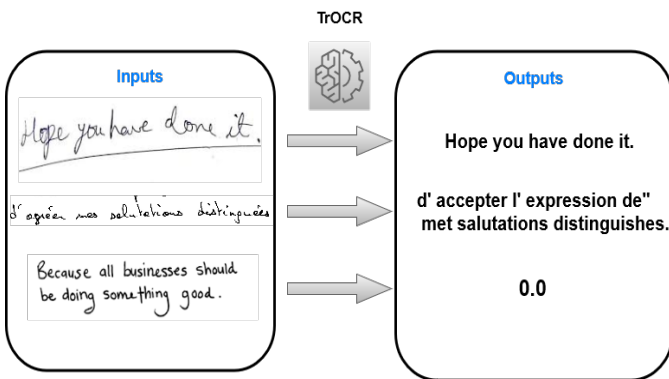


Fig. 2. TrOCR applied to three different examples

Figure 2 presents examples of TrOCR applications on simple handwritten text. The first row corresponds to a handwritten sentence in English. The model successfully digitized

it 100%. The second row relates to a handwritten sentence in French. We can see that the model managed to process it, but the digitization contains various errors, highlighting the model's limitation in **(1) processing French documents**. Finally, the third row corresponds to an attempt at a paragraph of three lines of text. The model could not process it and returned a null result, "0.0," illustrating the second limitation of the model when it comes to **(2) handling blocks of text**.

Although the performance of TrOCR was not optimal for handling French handwritten text and paragraphs, fine-tuning it on French data, along with a segmentation method, will help us overcome these two main issues. The entire approach will be detailed in the subsequent sections.

III. DIGITIZATION METHODOLOGY

To digitize EM reports, we selected the TrOCR model due to its efficiency in handwritten recognition. Additionally, it has an open-source version developed by Microsoft that allows easy fine-tuning, making it well-suited to our task. Additionally, an image segmentation method was developed to ensure that the input for our model consists of single-line text images, overcoming TrOCR's limitation in handling long blocks of text. We present, in Figure 3, our proposed methodology for digitizing medication error reports.

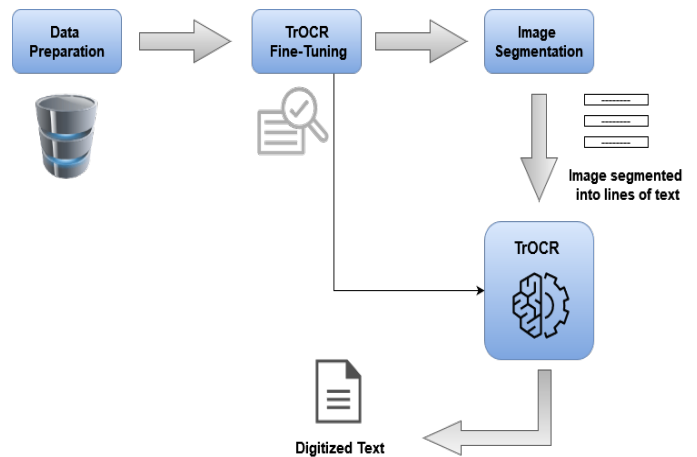


Fig. 3. Process followed for the digitization of ME

A. Data Preparation

In the training process of a DL model, it is necessary to properly prepare the data. This step involves applying image processing techniques to our dataset to eliminate noise and optimize image quality, thus facilitating feature distinction by the encoder. This step is followed by data labeling, which involves assigning each image the corresponding text, to better prepare the data for the training phase.

Initially trained on English data, the TrOCR model requires fine-tuning to improve its performance on French texts. To achieve this goal, we used the "RIMES 2011" dataset [15] for training the model. This dataset is a collection of over

10,000 images, each representing a unique sentence written in French and its corresponding text.

B. Fine-Tuning

For the training method, we chose sequence-to-sequence (Seq2Seq) training because of the length variability of the sequences of characters representing the input and output data. This type of training is generally used for NLP tasks such as machine translation, as well as other tasks like speech recognition and optical character recognition.

C. Image Segmentation

After preparing and adapting the TrOCR model, we chose to segment the initial images containing blocks of text representing the ME reports into sub-images, each containing a single line of text. Then, each segmented sub-image serves as input to the model sequentially, where OCR is applied to each line one at a time. This strategy allows us to overcome one of the constraints of the selected TrOCR model, namely its inability to process an entire paragraph at once. Figure 4 illustrates the different segmentation steps.

The segmentation of reports is performed according to the following sequence of steps:

- **Conversion of the image to grayscale and binarization:** In this step we will convert the RGB image into a grayscale image. Then, binarization is performed to represent the image in black and white pixels (binary image) to simplify upcoming processing.
- **Edge detection of the image:** In this step we will apply the Sobel filter [16]. It is an operator used in image processing to detect the edges in images. It involves identifying the text in the grayscale and binarized image.
- **Determination of the horizontal projection profile (HPP):** In this step, we will try to determine the exact position of the text in the input image to facilitate the segmentation process. One of the common methods to determine the interline spacing of a document is to analyze its horizontal projection profile (HPP). It is a 2D array representing the sum of the pixels in each line of the image. The HPP is a histogram where peaks correspond to lines of text, and troughs correspond to spaces between them. This informs us of the exact position of the lines of text in the image.
- **Text separation:** In this step we will detect peaks in the HPP and subdividing regions potentially associated with segments of lines in the text.
- **Application of the A* algorithm:** In this step, we will draw lines in the white space between each line of text to separate them within the image. To achieve this we will be using the A* (A Star) algorithm which is a path-finding algorithm in a graph between initial and final nodes. It uses heuristic evaluation on each node to estimate the optimal best path [17]. This algorithm is used to search for paths between the lines to accurately distinguish the position of each line and record the paths.

- **Drawing segmentation lines on the images:** In this step we will graphically represent the paths generated by the A* algorithm in the initial image.
- **Image segmentation :** In this step we will perform the segmentation based on the path lines resulting from the previous step.

IV. RESULTS

To digitize an ME report, we applied the previously detailed segmentation method and tested it on different reports. Figure 5 illustrates the digitization of a handwritten ME report with average handwriting where blue squares represent wrong transcriptions. For the preparation of this report, we performed manual processing to remove the lines drawn in the report and all the marks present on the report form that could affect both the segmentation and text generation. The accuracy of the predicted results is not quite efficient while it reached 40% correct words out of total words in the paragraph. On the other hand, Figure 6 shows the result of digitizing a handwritten report in handwriting text with beautiful calligraphy where blue squares represent wrong transcriptions. The accuracy of the predicted text was 89% which looks promising when it comes to a good calligraphy text.

We notice in Figure 5 that although the segmentation phase produced good results, the text generation phase needs improvement. On the other hand, the second attempt in Figure 6 shows that the digitization of a statement written in school handwriting, which has refined calligraphy, was entirely successful, including the text generation, despite the result having some more acceptable errors compared to the first test.

Fine-tuning the TrOCR model on French handwritten data and integrating a line-level segmentation technique effectively addressed the two main limitations of TrOCR for our task. Although the fine-tuning and segmentation phases were not flawless, the evaluation progress shows promising results for digitizing medication error reports.

V. CONCLUSION

This work proposes a novel architecture to digitize handwritten medication error reports. The proposed approach involves Deep Learning (DL) techniques. We presented its implementation by proposing a digitization process composed of several steps, including segmentation and model training. However, we encountered several obstacles, such as the particularity of the quality of the handwriting reports, which led to some failures in the segmentation and the text generation phases, concluded with not fully satisfactory results. To improve results, our future work will address these major challenges by attempting to refine the segmentation method further or even finding an alternative method to achieve greater success. Additionally, we will explore and adapt other OCR models to digitize the handwritten medication error reports with the highest possible quality. These improvements will be the subject of future publications.

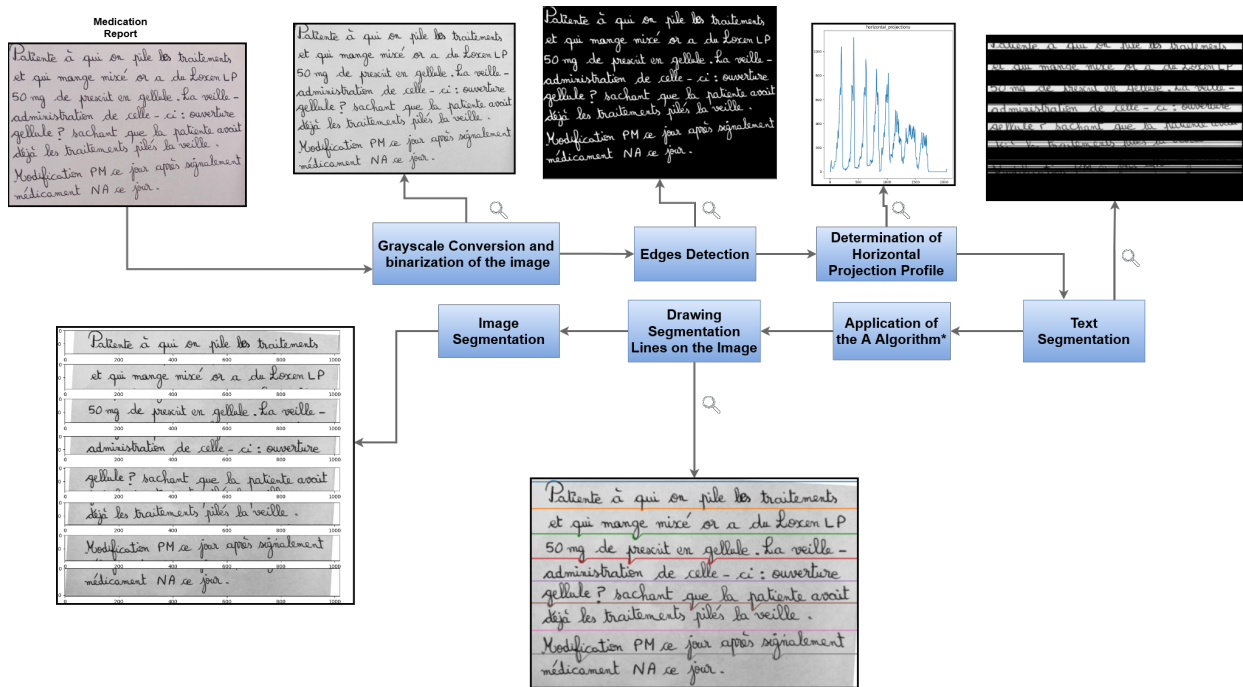


Fig. 4. Image Segmentation Process

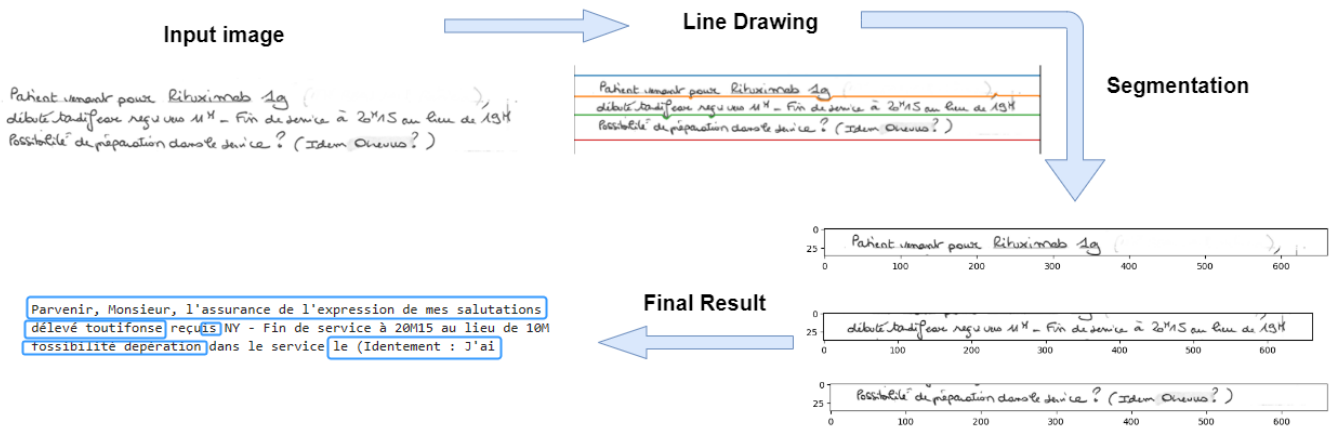


Fig. 5. Digitization of a Handwritten ME report with Average Handwriting

ACKNOWLEDGMENT

The authors would like to express their heartfelt appreciation to the Intercommunal Hospital Center of Castres-Mazamet and its esteemed medical professionals. Their willingness to provide essential information and insights is indispensable for the successful execution of this study. We are sincerely grateful for their willingness to share their knowledge and expertise, which played a pivotal role in shaping and advancing this research. Their feedback and insights have proven to be invaluable in guiding our work.

REFERENCES

[1] M. Abbasnazar, A. H. Talasaz, A. Eshraghi, and Z. Sahraei, "Detection and management of medication errors in internal wards of a teaching

hospital by clinical pharmacists," *Acta Medica Iranica*, pp. 482–486, 2013.

[2] H. Touati, R. Thabet, F. Fontanili, and E. Lamine, "Towards a digital collaborative framework for an efficient medication errors management," in *Working Conference on Virtual Enterprises*. Springer, 2023, pp. 549–562.

[3] R. Thabet, "Ingénierie dirigée par les modèles d'un pilotage robuste de la prise en charge médicamenteuse," Ph.D. dissertation, Institut National Polytechnique de Toulouse-INPT; Institut supérieur d . . . , 2020.

[4] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten optical character recognition (ocr): A comprehensive systematic literature review (slr)," *IEEE access*, vol. 8, pp. 142 642–142 668, 2020.

[5] Y. Lu, "Artificial intelligence: a survey on evolution, models, applications and future trends," *Journal of Management Analytics*, vol. 6, no. 1, pp. 1–29, 2019.

[6] D. Coquenot, C. Chatelain, and T. Paquet, "End-to-end handwritten paragraph text recognition using a vertical attention network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45,

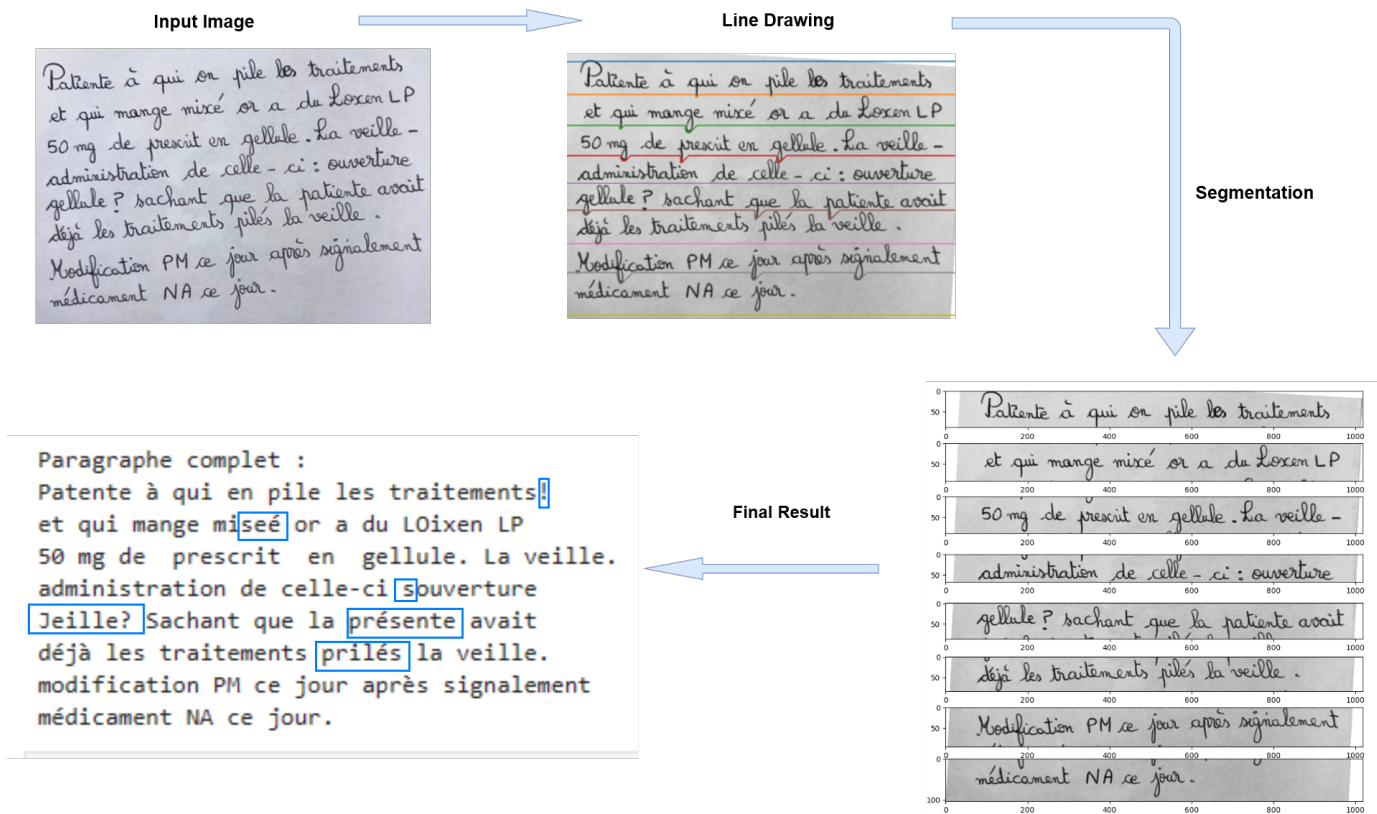


Fig. 6. Digitization of a Handwritten ME report with Good Handwriting

- no. 1, pp. 508–524, 2022.
- [7] —, “Dan: a segmentation-free document attention network for handwritten document recognition,” *IEEE transactions on pattern analysis and machine intelligence*, 2023.
 - [8] A. A. Idris and D. B. Taha, “Handwritten text recognition using crnn,” in *2022 8th International Conference on Contemporary Information Technology and Mathematics (ICITM)*. IEEE, 2022, pp. 329–334.
 - [9] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, “Trocr: Transformer-based optical character recognition with pre-trained models,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 13 094–13 102.
 - [10] M. Fujitake, “Dtroc: Decoder-only transformer for optical character recognition,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 8025–8035.
 - [11] D. H. Diaz, S. Qin, R. Ingle, Y. Fujii, and A. Bissacco, “Rethinking text line recognition models,” *arXiv preprint arXiv:2104.07787*, 2021.
 - [12] N. A. M. Isheawy and H. Hasan, “Optical character recognition (ocr) system,” *IOSR Journal of Computer Engineering (IOSR-JCE)*, e-ISSN, pp. 2278–0661, 2015.
 - [13] U.-V. Marti and H. Bunke, “The iam-database: an english sentence database for offline handwriting recognition,” *International Journal on Document Analysis and Recognition*, vol. 5, pp. 39–46, 2002.
 - [14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
 - [15] E. Grosicki¹, M. Carre, J.-M. Brodin, and E. Geoffrois¹, “Rimes evaluation campaign for handwritten mail processing,” 2008.
 - [16] O. R. Vincent, O. Folorunso *et al.*, “A descriptive algorithm for sobel image edge detection,” in *Proceedings of informing science & IT education conference (InSITE)*, vol. 40, 2009, pp. 97–107.
 - [17] Z. Zhang and Z. Zhao, “A multiple mobile robots path planning algorithm based on a-star and dijkstra algorithm,” *International Journal of Smart Home*, vol. 8, no. 3, pp. 75–86, 2014.