Architecture basée sur des modèles d'intelligence artificielle pour la numérisation de déclarations manuscrites des erreurs médicamenteuses

Brini Mohamed Ayachi[1,2], Touati Hanae [1], Thabet Rafika [1,3], Franck Fontanili [1], Lamine Elyes [1,4]

[1] CGI - IMT Mines Albi, Université de Toulouse, France {mohamed_ayachi.brini, hanae.touati , rafika.thabet, franck.fontanili } @mines-albi.fr, n° de téléphone : +335633900

[2] ISITCom, Université de Sousse, Tunisie

[3] Laboratoire MARS - LR17ES05, ISITCom, Université de Sousse, Tunisie

[4] ISIS, Institut National Universitaire Champollion, Université de Toulouse, France, elyes.lamine@univ-jfc.fr

Résumé. La reconnaissance optique de caractères (OCR) s'avère extrêmement utile dans plusieurs secteurs pour l'exploration des données massives archivées. Cette technologie permet la numérisation des textes imprimés ainsi que des textes manuscrits, qui sont fréquemment présents dans le domaine médical. Comme beaucoup de documents médicaux étaient jusqu'à présent rédigés manuellement, cela a entraîné l'accumulation de nombreuses données manuscrites, mais malheureusement impossible à exploiter. Leur numérisation par OCR est l'étape initiale à réaliser avant d'extraire automatiquement les données importantes de ces documents. Par exemple, dans les hôpitaux, la numérisation des déclarations manuscrites des erreurs médicamenteuses passées permet de peupler la base de données utile à l'implémentation de techniques d'analyse approfondie. Une telle base contribue à l'optimisation du processus de gestion des erreurs médicamenteuses. Pour accomplir cette numérisation de documents manuscrits, les algorithmes de Deep Learning (DL) ont démontré d'excellents résultats dans plusieurs études, bien que leur entraînement soit très coûteux en termes de temps et de ressources. Cependant, cet article détaille l'utilisation de modèles "Transformers" pré-entraînés, comme alternative moins coûteuse et qui donne des résultats plus conformes au texte manuscrit.

Mots clés : Reconnaissance Optique de Caractères (OCR), **Deep Learning (DL), Transformer**, **Gestion des erreurs médicamenteuses.**

Resumen. Espagnol

La tecnología de Reconocimiento Óptico de Caracteres (OCR) ha demostrado ser extremadamente útil en varios sectores para explorar grandes volúmenes de datos archivados. Esta tecnología permite la digitalización de textos impresos así como de textos manuscritos, que son comúnmente encontrados en el campo médico. Dado que muchos documentos médicos hasta ahora han sido escritos a mano, esto ha resultado en la acumulación de grandes cantidades de datos manuscritos, los cuales, lamentablemente, han sido imposibles de aprovechar. La digitalización mediante OCR es el primer paso necesario antes de poder extraer automáticamente datos importantes de estos documentos. Por ejemplo, en los hospitales, la digitalización de declaraciones manuscritas de errores de medicación pasados permite poblar una base de datos útil para la implementación de técnicas de análisis profundo. Tal base de datos contribuye a la optimización del proceso de gestión de errores de medicación. Para realizar esta digitalización de documentos manuscritos, los algoritmos de Aprendizaje Profundo (Deep Learning, DL) han demostrado excelentes resultados en varios estudios, aunque su entrenamiento es muy costoso en términos de tiempo y recursos. Sin embargo, este artículo detalla el uso de modelos "Transformers" preentrenados, como una alternativa menos costosa y que produce resultados más fieles al texto manuscrito.

Palabras claves : Reconocimiento Óptico de Caracteres (OCR), Aprendizaje Profundo (DL), Transformer, Gestión de Errores de Medicación.

Introduction

La sécurité des patients est une priorité absolue. Dans les établissements de santé et les hôpitaux, la gestion des risques et la prévention des incidents est essentielle pour l'amélioration continue de la qualité des soins. Parmi les nombreuses préoccupations, la gestion des erreurs médicamenteuses se présente comme un domaine crucial qui nécessite une attention immédiate. Les Erreurs Médicamenteuses (EM) sont définies comme chaque erreur non intentionnelle au cours du processus de soin impliquant un produit de santé ou bien un médicament, lors de la dispensation, l'administration ou bien la prescription [Le Beller et al., 2012] [Touati et al., 2023b]. Elles peuvent avoir des conséquences graves sur la santé des patients, ce qui justifie que leur prévention est essentielle pour réduire les risques d'incidents et améliorer l'efficacité des traitements médicaux.
Dans le cas de la gestion des EM, les professionnels de santé se basent sur des déclarations faites soit sur des fiches papier ou des formulaires numériques dédiés, afin d'analyser leurs causes et effets et proposer un plan d'action correctif et/ou préventif [Thabet, 2020] [Touati et al., 2023a]. Cependant, malgré l'émergence d'outils numériques dans la Prise en Charge Médicamenteuse, il existe encore de nombreuses déclarations d'EM manuscrites archivées et qui restent inexploitées. Ces documents comportent des informations de grande valeur sur les EM passées. S'ils étaient correctement numérisés et analysés, ils pourraient, en complément aux documents saisis dans des formulaires numériques, contribuer à l'optimisation du processus de gestion des erreurs médicamenteuses et à l'amélioration de la qualité et la sécurité des soins.
C'est pourquoi la technologie de reconnaissance optique de caractères (OCR) se révèle être une solution performante pour transformer les documents manuscrits en documents numériques. Cependant, l'application de l'OCR sur des données manuscrites présente plusieurs défis techniques et scientifiques, notamment en ce qui concerne le traitement de différentes écritures manuscrites, et la distinction entre l'écriture et différentes marques ou motifs pouvant apparaitre sur le papier, ce qui peut influencer la précision et la performance de la numérisation du texte manuscrit.
Il existe depuis longtemps de nombreuses techniques de reconnaissance optique de caractères. Dans cet article, nous nous intéressons principalement aux plus récentes à base d'intelligence artificielle. Les algorithmes de Deep Learning (DL) par exemple ont démontré d'excellents résultats sur l'extraction du texte à partir de données manuscrites, bien que leur entrainement soit très couteux en termes de temps et de ressources [Strubell et al. ,2019]. En alternative, ce papier présente une numérisation des archives des EM en utilisant un modèle "Transformer" ayant fait preuve d'excellents résultats dans différentes études publiées
Ce papier est structuré comme suit : Nous présentons dans la Section 2 les différentes techniques de reconnaissance optique de caractères, puis dans la Section 3, nous présentons notre cas d'application sur la numérisation des EM ainsi que l'architecture du modèle utilisée. Enfin, la Section 4 présente notre conclusion et nos perspectives de travaux futurs.

1. Les Techniques de reconnaissance optique de caractères (OCR)

1.1 La reconnaissance optique de caractères à partir de textes manuscrits

La reconnaissance optique de caractères OCR est un processus qui permet de lire un texte imprimé et le convertir en un format exploitable par l'ordinateur, tel que les codes ASCII. Grâce à un système OCR, les utilisateurs peuvent numériser des articles de magazines, des reçus, des tickets imprimés, et les convertir en des fichiers manipulables par un logiciel de traitement de texte par exemple. Les systèmes avancés de reconnaissance optique de caractères sont capables de lire des textes dans une grande variété de polices d'écriture numérique [Isheawy et al., 2015]. Quant à l'application de l'OCR sur des documents manuscrits, il s'agit de convertir des documents écrits manuellement par différents rédacteurs en un texte numérique modifiable et lisible par

l'ordinateur. Malgré la performance des techniques modernes, l'extraction du texte manuscrits présente un défi difficile à cause de la variabilité de l'écriture manuelle et de la qualité d'image.
En outre, l'extraction du texte à partir des documents manuscrits constitue un verrou majeur dans le domaine de l'intelligence artificielle (IA) : la variation stylistique, la qualité d'image, ainsi que l'état du papier, sont des facteurs qui accroissent la complexité du traitement des données manuscrites archivées. Heureusement, il existe de nombreux modèles en IA qui ont démontré d'excellents résultats dans différentes études durant les dernières années. Nous présentons dans le Tableau 1 ci-dessous quelques modèles et leurs performances qui sont testé sur la base de données IAM [Mart. et al. , 2002] qui contient 13 353 images de lignes de texte manuscrites créées par 657 rédacteurs. Pour chaque architecture, les auteurs ont choisi d'appliquer une évaluation par taux d'erreur par caractère (CER) comme métrique principale de comparaison. Le TrOCR, basé sur l'architecture classique du Transformer (encodeur + décodeur) a démontré des performances impressionnantes avec un CER de 2,89% seulement. Ce modèle a ensuite été amélioré en un nouveau modèle DTrOCR par [Fujitake et al., 2024]. Ce dernier se distingue par une architecture qui s'écarte du Transformer traditionnel, pour se baser uniquement sur un décodeur et un modèle de langage (LM) qui permet de générer du texte. Ce modèle présente un CER de 2,38% qui en fait le plus performant de tous les modèles du tableau. De plus, l'architecture combinant Self-attention + CTC + LM [Diaz et al., 2021] a enregistré un CER de 2,75%. Cette architecture utilise un bloc "Self-attention" avec un modèle de langage (LM) qui améliore la génération du texte et la perte CTC (Connectionist Temporal Classification) qui consiste à gérer la différence entre la longueur de séquences en entre et en sorite, afin d'optimiser le processus d'entrainement. En contraste, le modèle VAN (Vertical Attention Network) présente l'avantage de pouvoir extraire du texte à partir d'un paragraphe entier, contrairement aux autres modèles limités à une seule ligne de texte. Par contre, il a l'inconvénient de présenter le CER le plus élevé, valant 4,45%, avec un taux d'erreur par mot (WER) de 14,55%, ce qui présente une limitation en termes de performance. Enfin, l'architecture basique CRNN [Idris et al., 2022], qui intègre des couches de CNN (Convolutional Neural Network) et de RNN (Recurrent Neural Network), affiche un CER de 9,18% supérieur aux autres modèles. De plus, le Tesseract OCR de Google, couramment utilisé pour des textes déjà traité par ordinateur mais capable de s'adapter à des écrits manuscrits, a présenté un CER de 10% suite à une série des tests.
Cet état de l'art et ce benchmark publié des modèles d'OCR nous incite à opter pour le modèle TrOCR afin de construire notre architecture. En effet, il affiche une performance élevée avec un faible taux d'erreur de caractères, et il est également le seul parmi les modèles présentés à être pré-entrainé, nous permettant ainsi d'économiser des ressources et du temps

Tableau 1 Evaluation des différents modèles et architectures avec la métrique Taux d'erreur de caractères (CER) sur le jeu de données IAM

Modèle	TrOCR [Li. et al., 2023]		DTrOCR [Fujitake et al.,2024]	Self-attention +CTC+LM [Diaz et al., 2021]	VAN [Coquenet et al., 2022]	CRNN [Idris et al., 2022]	Tesseract OCR [Google]
Métrique	CER %	2,89%	2,38%	2,75%	4,45%	9,18 %	10%
	WER %				14,55%		

1.2 Modèle de reconnaissance optique de caractères basé sur les « Transformers »

Le modèle TrOCR [Li. et al., 2023] est basé sur l'architecture Transformer comme présenté dans la Figure 2. Il est composé d'un Transformer d'image chargé de l'extraction des caractéristiques visuelles et d'un Transformer de texte en charge de la modélisation linguistique. La structure adoptée par le TrOCR est la structure encodeur-décodeur du modèle Transformer classique. L'encodeur est conçu pour capturer la représentation des segments d'image, tandis que le décodeur vise à générer la séquence de morceaux de mots, guidée par les caractéristiques

visuelles et les prédictions précédentes. L'encodeur et le décodeur sont équipés d'un mécanisme d'attention qui permet de calculer pour chaque mot un score déterminant la position du mot dans la phrase générée afin d'améliorer la qualité du texte en sortie et le rendre plus compréhensible.

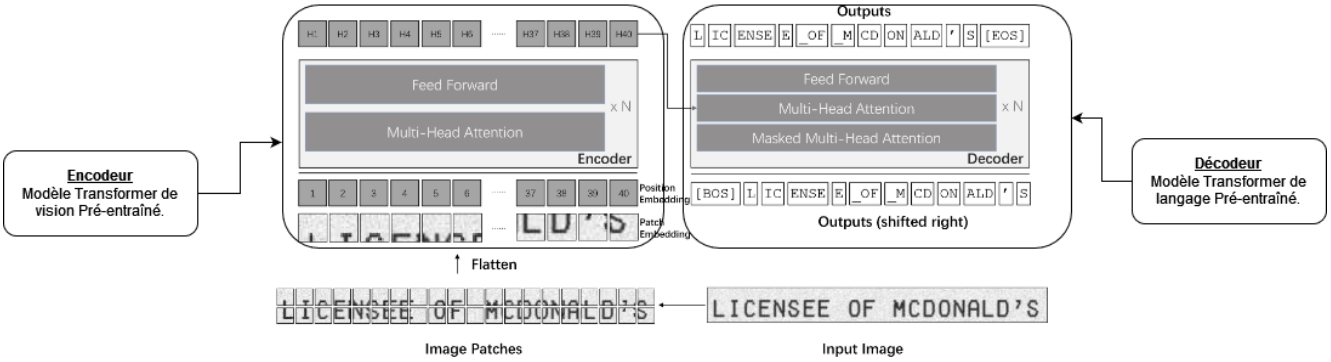
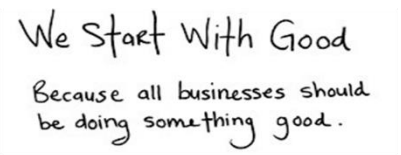


Figure 1 Architecture du modèle TrOCR [Li. et al. , 2023]

Le TrOCR est un modèle pré-entraîné qui est donc plus avantageux en termes de réduction du temps et des ressources nécessaires. Il offre une précision sur la prédiction améliorée grâce au mécanisme d'attention intégré dans son architecture de Transformer classique. Cependant, comme il a été entrainé sur des textes en langue anglaise, il est moins performant en langue française. Par ailleurs, l'efficacité de ce modèle repose sur la condition que chaque image soumise doit contenir uniquement une ligne de texte, nécessitant ainsi une segmentation initiale de l'image en plusieurs sous-images. Chacune de ces dernières doit inclure une unique ligne de texte pour garantir un traitement optimisé. Dans de nombreuses études précédentes [Conquenet et al., 2022], [Conquenet et al., 2023], au lieu de segmenter l'image en sous-images pour isoler chaque ligne de texe dans un paragraphe, les auteurs ont utilisé des architectures du Deep Learning pour identifier les lignes au sein des paragraphes. Un exemple pertinent est le modèle VAN [Conquenet et al., 2022] qui emploie un bloc FCN (Fully Convolutional Network) et qui est composé de plusieurs blocs CNN pour effectuer cette tâche. Toutefois, bien que cette approche soit efficace, elle nécessite des ressources en terme du temps et de matériel pour entrainer l'architecture a distinguer les lignes dans chaque paragraphe.

Tableau 2 TrOCR appliqué à trois exemples différents

Entrées	TrOCR	Sorties
		'of the House of Lords, but while it remains'
		'Envous emerciont d'avance de we le faire'
		'0 0'

Le Tableau 2 présente des exemples d'applications du TrOCR sur du texte manuscrit simple. La première ligne correspond à une phrase manuscrite en anglais. Le modèle a réussi à la numériser à 100%. La deuxième ligne concerne une phrase manuscrite en français. On voit que le modèle a réussi à la traiter, mais la numérisation comporte différentes erreurs, soulignant les limitations du modèle sur le traitement des documents

en français. Finalement, la troisième ligne du tableau 2 correspond à un essai fait sur un paragraphe de trois lignes de texte. Le modèle n'a pas pu le traiter et donne un résultat nul en sortie « 0.0 », ce qui illustre la deuxième limitation du modèle concernant son incapacité à traiter des blocs de texte.

2 Numérisation des déclarations des EM avec du TrOCR

Pour la numérisation des déclarations des EM avec le modèle TrOCR, nous proposons de suivre l'approche de la Figure 2 selon les étapes qui suivent.

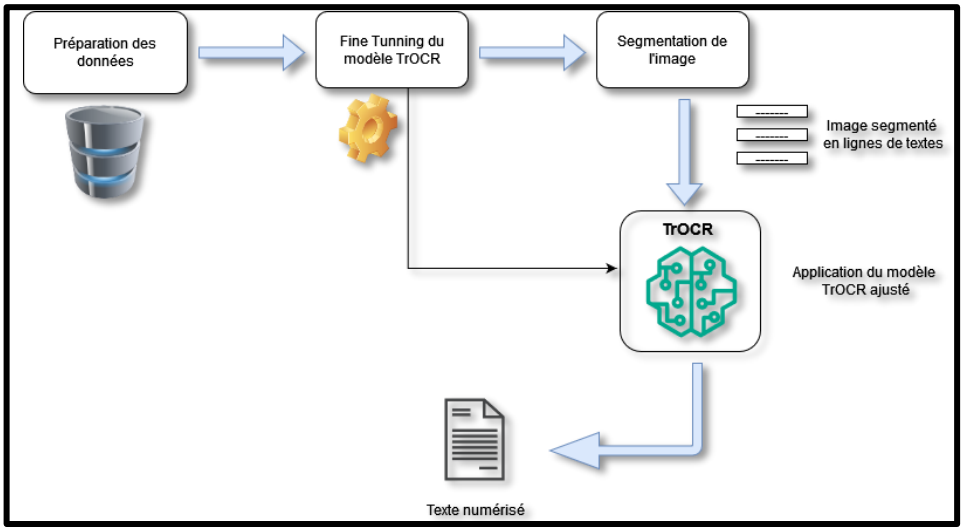


Figure 2 Processus suivi pour la numérisation des EM

2.1 Préparation des données et adaptation du modèle TrOCR

Dans un processus d'entraînement d'un modèle de Deep Learning (DL), il est préalable de bien préparer ses données. Cette étape implique l'application de techniques de traitement d'images à notre jeu de données pour éliminer le bruit et optimiser la qualité des images, facilitant ainsi la distinction des caractéristiques par l'encodeur. Cette étape est suivie par l'étiquetage des données qui consiste à donner pour chaque image un nom et le texte qui lui correspond, ceci dans le but de mieux préparer les données pour l'étape d'entrainement.
Initialement entraîné sur des données en anglais, le modèle TrOCR nécessite un ajustement fin appelé « Fine Tuning » pour améliorer sa performance sur des textes en français. Afin d'atteindre cet objectif, nous avons utilisé le jeu de données « RIMES 2011 » [Grosicki et al., 2008] pour l'entraînement du modèle. Il s'agit d'une collection de plus de 10000 images, dans laquelle chaque image représente une phrase unique rédigée français avec un ensemble des textes qui correspond à chacune des phrases. Nous avons choisi l'entrainement à séquence « Seq2Seq », car les données en entrée et celles en sortie sont des séquences de caractères de longueur variable. Ce type d'entrainement est généralement utilisé pour les taches du NLP comme la traduction automatique, et pour d'autres tâches de reconnaissance vocale et de reconnaissance optique de caractères.

2.2 Segmentation des images

Après la préparation et l'adaptation du modèle TrOCR, nous avons choisi de segmenter les images initiales avec un bloc de texte représentant les déclarations des EM en des sous-images, chacune renfermant une unique ligne de texte. Ensuite, chaque sous-image segmentée sert d'entrée du modèle d'une façon séquentielle et sur laquelle

nous appliquerons l'OCR pour chaque ligne à la fois. Cette stratégie permet de contourner l'une des contraintes du modèle TrOCR sélectionné, à savoir son incapacité à traiter un paragraphe entier en une seule fois. La figure 2 illustre les différentes étapes de la segmentation.

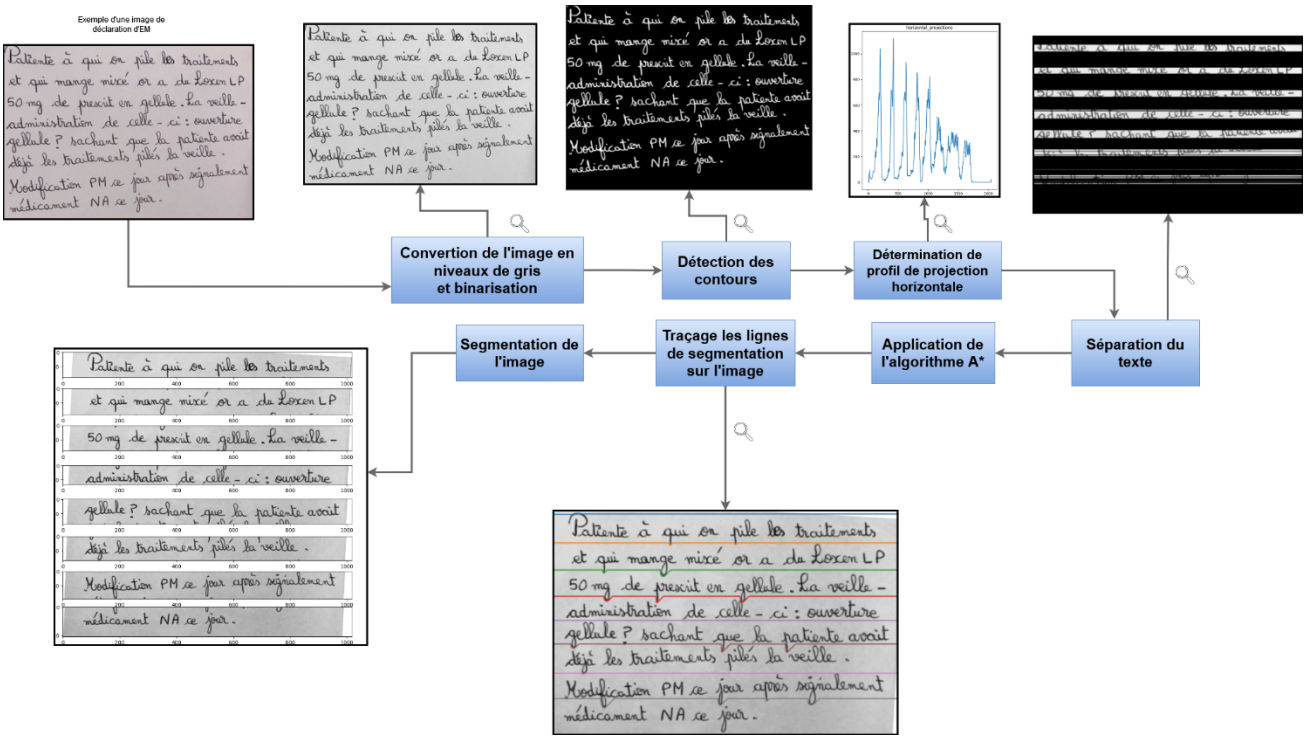


Figure 3 Démarche suivie pour la segmentation des images

La segmentation des déclarations se fait selon la succession d'étapes suivantes :

- **Conversion de l'image en niveaux de gris et binarisation :** Cette étape consiste à convertir l'image en couleur en une image en niveaux de gris. Ensuite, nous passons à la binarisation pour représenter l'image en des pixels noir et blanc (image binaire) afin de simplifier les traitements ultérieurs.
- **Détection des contours d'image :** Cette étape repose sur l'application du filtre de Sobel [Vincent. et al., 2009]. Il s'agit d'un opérateur utilisé en traitement d'image pour détecter les contours des images. Il consiste à repérer le texte dans l'image qui a été mis en niveaux de gris et binarisé.
- **Détermination du profil de projection horizontale (HPP) :** L'une des méthodes courantes pour déterminer l'interligne d'un document consiste à analyser son profil de projection horizontal (HPP). Il s'agit d'un tableau en 2 dimensions qui représente la somme des pixels dans chaque ligne de l'image. Le HPP est un histogramme où les pics correspondent aux lignes de texte et les creux correspondent aux espaces entres elles. Cela nous informe de la position exacte des lignes du texte dans l'image.
- **Séparation du texte :** Cette étape consiste à détecter les pics dans le HPP et la subdivision des régions potentiellement associées à des segments des lignes dans le texte.
- **Application de l'algorithme A* :** L'algorithme A* (A Star) est un algorithme de recherche de chemin dans un graphe entre un nœud initial et un nœud final. Il utilise une évaluation heuristique sur chaque nœud pour estimer le meilleur chemin optimal [Zhang Et al., 2014]. Cet algorithme est utilisé pour la recherche des trajectoires entres les lignes pour bien distinguer la position de chaque ligne et enregistrer les trajets.
- **Traçage des lignes de segmentations sur les images :** Cette étape consiste à représenter graphiquement les trajectoires générées par l'Algorithme A* dans l'image initiale.

- **Segmentation de l'image :** Cette étape consiste à réaliser la segmentation en se basant sur les lignes des trajectoires résultant de l'étape précédente.

2.3 Application à la numérisation d'une déclaration d'EM

Pour réaliser la numérisation d'une déclaration d'EM, nous avons appliqué la segmentation précédemment détaillée que nous avons testée sur différentes déclarations. La figure 4 illustre la numérisation d'une déclaration d'EM manuscrite avec une calligraphie assez moyenne. La figure 5 montre le résultat de la numérisation d'une déclaration manuscrite en écriture écolière avec une très belle calligraphie. Pour la préparation de la première déclaration, nous avons réalisé des traitements manuels permettant d'éliminer les lignes tracées dans la déclaration et toutes les marques "parasites" présentes sur la fiche de déclaration qui peuvent affecter à la fois la segmentation et la génération du texte.

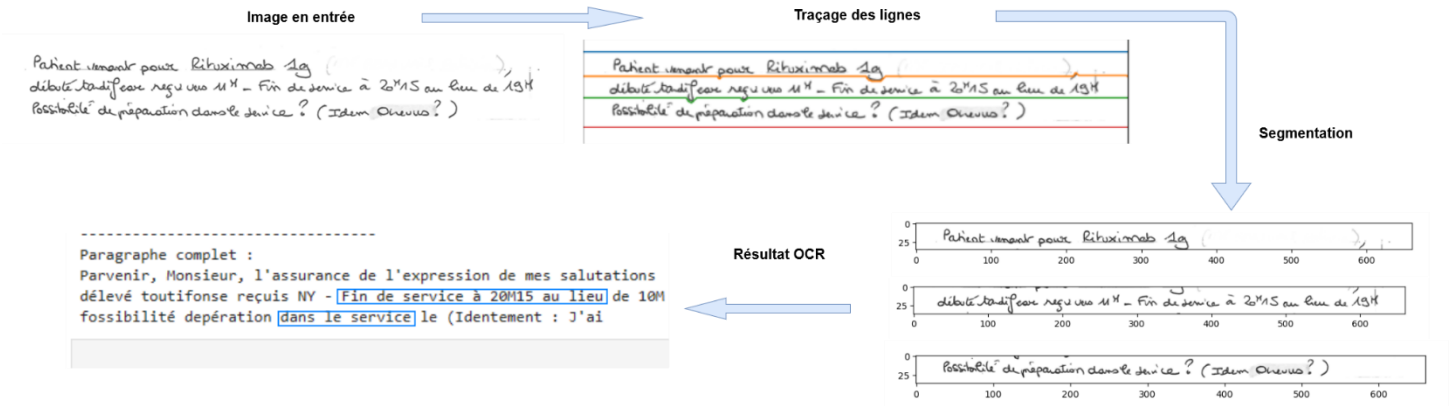


Figure 4 Numérisation d'une déclaration d'EM manuscrite avec une calligraphie moyenne

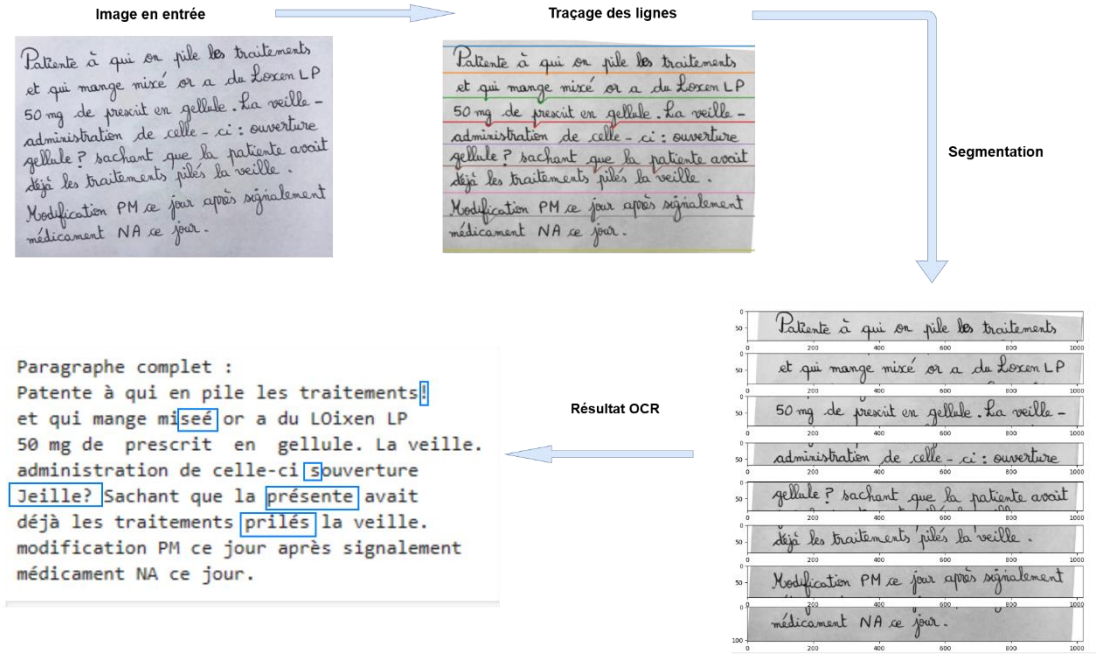


Figure 5 Numérisation d'une déclaration d'EM manuscrite avec une bonne calligraphie

Nous remarquons sur la Figure 4 que bien que la phase de segmentation ait donné de bons résultats, la phase de génération du texte nécessite une amélioration. D'autre part, le deuxième essai de la Figure 5, montre que la numérisation d'une déclaration rédigée en écriture écolière, à savoir une écriture présentant une calligraphie soignée, a été entièrement réussie, y compris pour la génération du texte, malgré que le résultat présente quelques erreurs plus acceptables par rapport au premier test.

Conclusion

Dans ce papier, nous avons présenté l'avancement de nos premiers travaux destinés à la numérisation des déclarations des erreurs médicamenteuses manuscrites. L'approche que nous avons utilisée fait appel à des techniques de Deep Learning. Nous avons présenté sa mise en œuvre en proposant un processus de numérisation composé de plusieurs étapes, notamment la segmentation et l'entrainement du modèle. Cependant, les résultats ne sont pas encore pleinement satisfaisants. Plusieurs obstacles ont causé des échecs au niveau de la segmentation et aussi dans la phase de génération du texte, notamment la mauvaise calligraphie de l'écriture des déclarations. Nos prochains travaux vont donc s'attacher à lutter contre ces difficultés majeures en tentant de mieux affiner la méthode de segmentation, ou même de trouver une méthode alternative pour y parvenir avec plus de succès. De plus, nous devons encore explorer et adapter d'autres modèles d'OCR afin d'arriver à numériser des déclarations d'erreurs médicamenteuses manuscrites avec la meilleure qualité possible.

Références

Coquenet, D., Chatelain, C. and Paquet, T., 2022. End-to-end handwritten paragraph text recognition using a vertical attention network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), pp.508-524.

Coquenet, D., Chatelain, C. and Paquet, T., 2023. DAN: a segmentation-free document attention network for handwritten document recognition. *IEEE transactions on pattern analysis and machine intelligence*.

Diaz, D.H., Qin, S., Ingle, R., Fujii, Y. and Bissacco, A., 2021. Rethinking text line recognition models. *arXiv preprint arXiv:2104.07787*.

Fujitake, M., 2024. Dtrocr: Decoder-only transformer for optical character recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 8025-8035).

Grosicki[1], E., Carre, M., Brodin, J.M. and Geoffrois[1], E., 2008. RIMES evaluation campaign for handwritten mail processing.

Isheawy, N.A.M. and Hasan, H., 2015. Optical character recognition (ocr) system. *IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN*, pp.2278-0661. »

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z. and Wei, F., 2023, June. Trocr: Transformer-based optical character recognition with pre-trained models. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 11, pp. 13094-13102

Marti, U.V. and Bunke, H., 2002. The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, *5*, pp.39-46

Thabet, R., 2020. *Ingénierie dirigée par les modèles d'un pilotage robuste de la prise en charge médicamenteuse* (Doctoral dissertation, Institut National Polytechnique de Toulouse-INPT ; Institut Supérieur d'Informatique et des Techniques de Communication-ISITCom (Sousse, Tunisie)).

Touati, H., Thabet, R., Fontanili, F. & Lamine, E. (2023a), Towards a digital collaborative framework for an efficient medication errors management, in *'Working Conference on Virtual Enterprises'*, Springer, pp. 549–562

Touati, H., Thabet, R., Fontanili, F., Pingaud, H., Cleostrate, M.-H., Cufi, M.-N., Pruski, M. & Lamine, E. (2023b), Vers une démarche sécurisée des erreurs médicamenteuses à l'hôpital, in *'CIGI QUALITA MOSIM 2023'*,pp.1–8

Vincent, O.R. and Folorunso, O., 2009, June. A descriptive algorithm for sobel image edge detection. In *Proceedings of informing science & IT education conference (InSITE)* (Vol. 40, pp. 97-107).

Zhang, Z. and Zhao, Z., 2014. A multiple mobile robots path planning algorithm based on A-star and Dijkstra algorithm. *International Journal of Smart Home*, *8*(3), pp.75-86.