Handwritten Character Recognition

RAMDANI Nabil[1], IBEN JELLAL Hamza[2], SAMI Youssef[3], TAMIM Yassine[4], EL HIZAZ Salwa[5] and AHIDAR Adil,[a]

[1] *Ecole Centrale Casablanca*
[2] *Ecole Centrale Casablanca*
[3] *Ecole Centrale Casablanca*
[4] *Ecole Centrale Casablanca*
[5] *Ecole Centrale Casablanca*
[a] *Supervisor*

Reception date of the manuscript: 10/01/2024
Acceptance date of the manuscript: 10/01/2024
Publication date: 10/01/2024

Abstract— Optical Character Recognition (OCR) technology holds significant potential for digitizing handwritten and printed text, particularly in the healthcare sector for processing medical prescriptions. This study outlines a comprehensive approach to OCR system development, beginning with the creation of a character dataset that encompasses various typefaces. Our initial model, trained on this dataset, demonstrated promising results when tested on individual characters. However, when applied to medical prescriptions, it yielded suboptimal outcomes. Consequently, we refined our approach by maintaining the original model architecture and retraining it on an alternative dataset, specifically the "EMNIST" dataset. This adjustment led to substantial improvements, achieving a CNN accuracy of 86%.

Keywords— Optical Character Recognition (OCR), Deep Learning, Convolutional Neural Networks (CNN), Character Recognition

I. Introduction

In In the dynamic landscape of information technology, the intersection between handwriting recognition and machine-generated text synthesis emerges as a critical domain, presenting challenges in the interface between human-written communication and machine interpretation. This introduction sets the stage for a systematic exploration, commencing with a contextualization of the theoretical framework, followed by a nuanced problematization of the research question, formulation of research hypotheses, and a brief overview of the planned scientific inquiry.

The theoretical framework initiates an examination of the current state of handwriting recognition and machine-generated text synthesis, elucidating key terms and concepts fundamental to the field. This foundation is crucial for understanding the intricate dynamics involved in translating human-written text into machine-generated content. Within this theoretical realm, emphasis is placed on the integration of convolutional neural networks (CNN) and optical character recognition (OCR), providing a structured pathway for a comprehensive analysis of the conversion process from handwritten text to machine-generated text.

Contact data: IBEN JELLAL Hamza, hamza.ibenjellal@gmail.com

As the investigation progresses, focus shifts towards the complexities inherent in translating handwritten text into machine-generated content. Beyond technological intricacies, the challenge extends to bridging the subtle gap between human written expression and machine interpretation. Central questions revolve around determining the optimal approach for seamless transfer while identifying and addressing potential disparities and limitations arising in this intricate process. This section aims to unravel the inherent complexities of transforming human written language into machine-understandable text.

Building upon the problematization, the research hypothesis positions convolutional neural networks (CNN) and optical character recognition (OCR) as pivotal elements for achieving a robust and efficient conversion of handwritten text into machine-generated content. The proposed approach seeks to harness the strengths of CNNs and OCR to surmount existing limitations, providing a pathway to bridge the gap between human-written and machine-generated language. As this hypothesis is developed, the subsequent section of the article is dedicated to empirically testing this proposed approach.

Anticipating the unfolding narrative, a structured plan has been devised to guide the scientific study. The subsequent sections commence with a comprehensive review of related work and literature, shedding light on existing solutions and their merits and shortcomings. The methods section then meticulously outlines the experimental procedures, incorporating the use of CNN and OCR, aiming to enhance reproducibility for fellow researchers. The "Results and Discussion" section subsequently unveils the outcomes of the experimental procedures, presenting results in a format conducive to comparison and analysis. Finally, the conclusion succinctly encapsulates this research, highlighting its implications and potential avenues for future exploration at the dynamic intersection of handwriting recognition and machine-generated text synthesis.

II. STATE OF THE ART:

a. Convolutional Neural Networks (CNN)

Architecture Design

Recent advancements in Convolutional Neural Networks (CNNs) for handwritten recognition emphasize the development of deeper and more intricate structures. For instance, the integration of residual connections in ResNet architectures addresses the vanishing gradient problem, facilitating the training of exceptionally deep networks. Moreover, attention mechanisms, as observed in Transformer-based models, play a pivotal role by allowing the network to concentrate on specific regions of input, proving crucial for recognizing diverse writing styles and intricate character formations.

Data Augmentation and Transfer Learning

To enhance the robustness of models, especially when faced with limited labeled data, employing data augmentation techniques such as rotation, scaling, and flipping is imperative. Additionally, transfer learning, involving the utilization of pre-trained models from extensive datasets like ImageNet, has proven beneficial. Fine-tuning these models specifically for handwritten recognition tasks amplifies performance, especially in scenarios with constrained datasets.

Regularization Techniques

Common regularization techniques, including Dropout and Batch Normalization, are crucial to prevent overfitting in deep networks. The application of these methods significantly contributes to the model's generalization capabilities when applied to handwritten recognition tasks.

b. Optical Character Recognition (OCR)

Preprocessing Techniques

Effective preprocessing techniques, such as adaptive thresholding and morphological operations, play a pivotal role in binarization and noise reduction in handwritten documents. Preprocessing is particularly critical in enhancing the signal-to-noise ratio, especially in documents with diverse backgrounds and varying quality.

RNNs and LSTMs

Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs) are preferred for recognizing sequences of characters in handwriting. Their ability to capture temporal dependencies between successive characters makes them particularly suitable for cursive and connected writing styles.

Connectionist Temporal Classification (CTC)

CTC emerges as a key technique for sequence-based recognition tasks, enabling the training of models without the need for explicit alignment between input and output sequences. This flexibility proves invaluable in handling variable-length sequences in handwritten text.

c. Feature Extraction Techniques

Handcrafted Features

Effective handcrafted features like Histograms of Oriented Gradients (HOG) and Local Binary Patterns (LBP) contribute significantly to capturing local gradient and texture information, respectively. These features are essential for distinguishing various character shapes in handwritten recognition.

Deep Feature Extraction

Convolutional layers in CNNs automatically learn hierarchical features. Initial layers capture low-level features like edges and corners, while deeper layers excel in capturing more complex patterns and structures. Spatial Pyramid Pooling (SPP) and Global Average Pooling (GAP) are techniques employed for capturing spatial information at different granular levels, enabling the model to recognize characters irrespective of their size or position in the input image.

d. Comparaison

CONVOLUTIONAL NEURAL NETWORKS (CNN):

- **Description:** CNNs are deep learning architectures designed for spatial hierarchies, widely used in image-based tasks.
- **Challenges:**
  - Automatically learn hierarchical features.
  - Effective in image recognition tasks.
  - Proneness to overfitting, especially with limited data.

OPTICAL CHARACTER RECOGNITION (OCR):

- **Description:** OCR involves the recognition of text, both printed and handwritten, often using machine learning techniques.
- **Challenges:**
  - Enables digitization of text from various sources.
  - Applicable to a wide range of documents.
  - Sensitivity to variations in writing styles and conditions.

FEATURE EXTRACTION TECHNIQUES:

- **Description:** Techniques involve extracting relevant features from raw data, crucial for model learning and recognition.
- **Challenges:**
  - Handcrafted features provide interpretability.
  - Deep learning automatically learns features.
  - Handcrafted features may not capture complex patterns.
  - Deep learning models may lack interpretability.

e. Recent Trends

End-to-End Learning Approaches

Integrating attention mechanisms in models like Transformer OCR streamlines the mapping process from input images to character sequences, eliminating the need for separate modules for feature extraction and sequence decoding. **Generative Adversarial Networks (GANs)**

The utilization of Generative Adversarial Networks (GANs) for generating synthetic handwritten data proves beneficial when obtaining a large amount of real labeled data is challenging. Synthetic data generated by GANs is instrumental in training models, particularly in scenarios with data scarcity.

III. Methodology

a. Convolutional Neural Networks (CNNs) in Handwriting Recognition: A Scientific Perspective

Convolutional Neural Networks (CNNs) are a class of deep neural networks, predominantly used in the field of computer vision. Their structure is inspired by the organization of the animal visual cortex and is particularly adept at processing data with a grid-like topology, such as images. The core idea of a CNN is the convolutional layer, which applies a convolution operation to the input data. Mathematically, this can be represented as:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

In the context of CNNs for image processing, this formula is discretized and simplified. For a 2D image I and a filter K, the convolution is:

$$(I * K)(i, j) = \sum_{m}\sum_{n} I(m,n)K(i-m, j-n)$$

where $I(m,n)$ is the pixel value at position (m,n) in the image, and $K(i-m, j-n)$ is the filter/kernel value at the corresponding position.

b. Application in Handwritten Character Recognition

In handwriting recognition, CNNs process an input image of handwritten text and learn to recognize and interpret individual characters and words. The variability in handwriting styles, sizes, and distortions presents a unique challenge, effectively addressed by CNNs.

A typical CNN architecture for this task would involve several convolutional layers, each detecting features at different levels of abstraction. For example, the first layer may detect edges and basic shapes, while deeper layers might identify more complex structures like letters.

1. Generation and Augmentation of a Character Dataset for Optical Character Recognition (OCR)

Original Dataset Creation Our dataset has been meticulously curated to encompass a wide array of characters, including digits from 0 to 9 and lowercase letters from a to z. We selected a diverse set of typefaces ranging from formal to more casual and handwritten styles, ensuring that our deep learning model could learn to recognize characters across varied contexts. Each character was rendered in black against a white background to create grayscale images of 50×50 pixels in size, which aligns with the input resolution required for our Convolutional Neural Network (CNN).

Dataset Augmentation Techniques To enhance our model's generalization capabilities, we applied data augmentation techniques to our original dataset. This augmentation was aimed at simulating natural variations and imperfections encountered in real-world scenarios. The following methods were employed: - Additive Noise: Gaussian noise was added to images to simulate the effects of ink variations and minor image disturbances. - Photometric Alterations: Adjustments to contrast and brightness were applied to replicate the effects of different lighting conditions and print qualities. These transformations were carefully applied to maintain character readability while introducing realistic diversity.

Data Visualization and Analysis To demonstrate the diversity of our dataset, we present below examples of characters from different typefaces and the variations introduced by data augmentation.

Fig. 1: examples of characters from our dataset

The visualizations reveal the richness of the dataset and the effectiveness of augmentation in creating a robust data set, ready to be used for training advanced character recognition models.

The following figure illustrates the class distribution within our dataset, highlighting the balanced representation of characters which is crucial for the unbiased training of our convolutional neural network model.

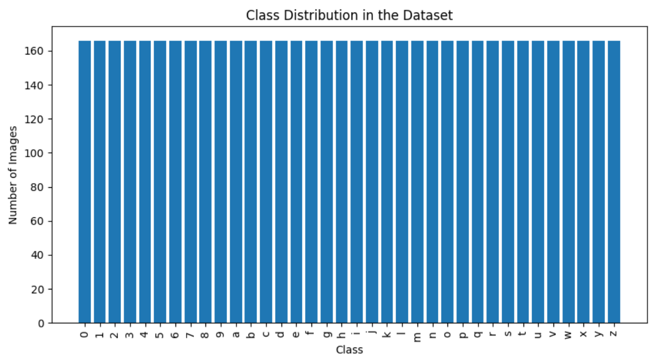

Fig. 2: Overview of the distribution of elements in our dataset

2. Convolutional Neural Network (CNN) Architecture for Character Recognition

Our study employed a Convolutional Neural Network (CNN), a class of deep neural networks highly effective for analyzing visual imagery. The architecture of our CNN was designed to recognize a mixed set of alphanumeric characters, encompassing both letters and digits. The model's input layer accepts 50×50 pixel grayscale images, reflecting the standardized preprocessing applied to our dataset.

CNN Layers and Their Functions

1. Convolutional Layers: The model comprises two convolutional layers, each followed by a max-pooling layer. The first convolutional layer has 32 filters of size 3×3 and employs ReLU activation to introduce non-linearity. The subsequent max-pooling layer reduces spatial dimensions by half, thus focusing on the most relevant features. The second convolutional layer doubles the filters to 64, continuing the pattern of capturing complex patterns within the image data.

2. Regularization: Each convolutional layer includes L2 regularization to mitigate overfitting by penalizing large weights, ensuring the model's generalizability to new data.

3. Flattening Layer: Post feature extraction, a flattening layer transforms the 2D matrix data into a 1D vector, making it suitable for input into the dense layers.

4. Dense Layers: Following flattening, a dense layer with 128 units further processes the features, employing ReLU activation. The final dense layer corresponds to the number of classes (36, representing digits and lowercase letters) and uses the softmax activation function to produce a probability distribution over the classes.

3. Training and Evaluation

Our CNN was trained using an Adam optimizer with a categorical crossentropy loss function. The model underwent 10 epochs of training, with validation data providing an insight into the model's performance on unseen data throughout the process. The test accuracy achieved was a significant $\sim$90%, demonstrating the effectiveness of the model in character recognition tasks.

The confusion matrix offers a detailed view of the model's performance across all classes, highlighting its strengths and areas where improvements could be made.

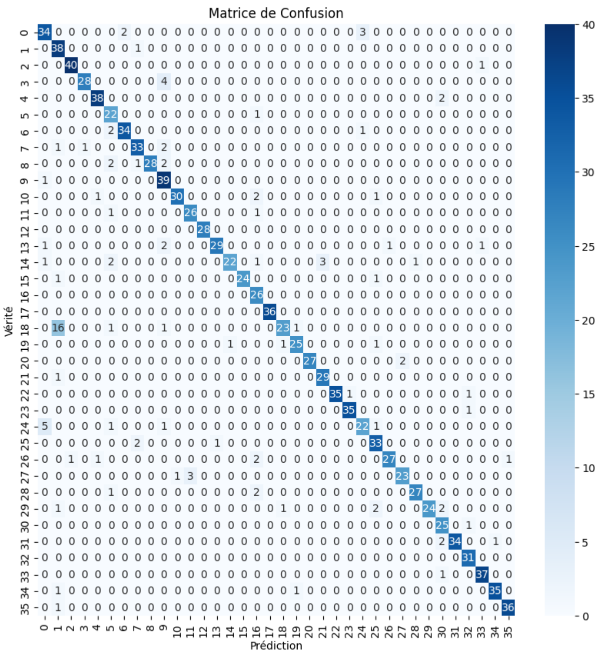

Fig. 3: Confusion Matrix

The confusion matrix above reveals a notable misclassification between the numeral '0' and the lowercase letter 'o', as evidenced by the off-diagonal entries corresponding to these classes.

To address this challenge, we have introduced contextual post-processing functions. These functions utilize surrounding character context to disambiguate between '0' and 'o'. For instance, when the ambiguous character 'X' (representing either '0' or 'o') is flanked by letters, it is classified as 'o'. Conversely, if 'X' is preceded or followed by a numeral, it is more accurately identified as '0'. This context-aware approach significantly reduces misclassifications and improves the overall accuracy of our text recognition pipeline.

Accompanying our model description, Figures 8 and 9 visually summarize the training process across epochs, showcasing the loss and accuracy metrics for both training and validation phases.

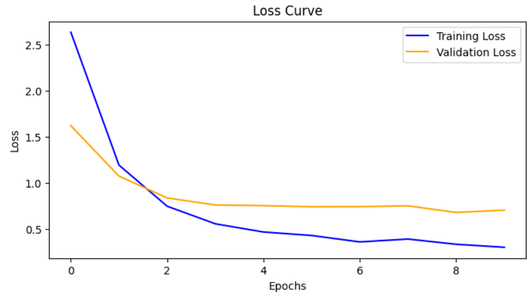

Fig. 4: Training and Validation Loss

Figure 8 depicts the loss curves, a fundamental indicator of the learning process. The training loss shows a steep decline, indicating that the model is effectively learning from the dataset. Concurrently, the validation loss decreases alongside the training loss, suggesting that the model is not memorizing the training data but generalizing well to unseen data. The convergence of both curves implies a consistent learning trend without significant overfitting.

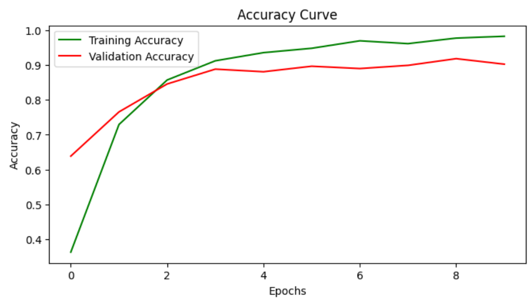

Fig. 5: Training and Validation Accuracy

Figure 9 presents the accuracy curves, reflecting the model's capacity to correctly classify the data. The training accuracy curve ascends sharply and continues to rise, indicating that the model increasingly learns to identify the correct characters. The validation accuracy initially follows the training accuracy closely but later plateaus, which might suggest that the model has reached its potential given the current architecture and data. It's noteworthy that the validation accuracy maintains a high level, which corroborates the model's robustness.

Both figures collectively demonstrate the model's competence in recognizing characters from various typefaces and under different data augmentation conditions. The performance metrics illustrated here reinforce the model's suitability for complex OCR tasks in diverse settings.

c. Application in Handwritten Medical Prescriptions Recognition

1. Generation of a Medical Prescription Dataset

To extend the practical application of our character recognition model, we generated a synthetic dataset of medical prescriptions. The prescriptions were created using the same diverse set of typefaces that we used for individual character generation, ensuring stylistic consistency and facilitating the transition from character-level to document-level recognition.

Prescription Generation Methodology

We employed a scripted approach to simulate the composition of medical prescriptions. This involved the random assembly of common medication names, dosages, frequencies, and additional instructions, closely mimicking the structure of authentic prescriptions. Each generated text string represented a realistic prescription directive, such as "Prendre 2 Ibuprofène 400mg toutes les 8 heures avec un grand verre d'eau."

Image Rendering

For each prescription text, we programmatically rendered the text onto a white canvas of 1000×500 pixels to simulate paper prescriptions. The choice of font for each prescription was randomized from the same collection of typefaces used in the character dataset to maintain a consistent challenge for the OCR system. This process ensured that the prescriptions varied not only in textual content but also in visual presentation due to the different typographic styles.

Dataset Composition

The final dataset comprised 100 unique prescription images. We captured the image paths and corresponding prescription texts in a CSV file, facilitating easy access and reference for training and evaluation purposes.

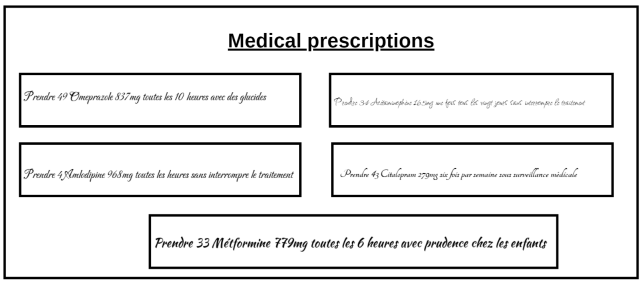

Fig. 6: Sample Medical Prescriptions from the Generated Dataset

2. Application of the Trained Model to Prescription Recognition

Building on our character recognition capabilities, we attempted to apply our trained CNN model to the task of prescription text recognition. Despite the success in character-level identification, the direct application of the model to prescription images yielded suboptimal results. The primary challenge lay in the character segmentation phase, a crucial step where a prescription image is divided into individual characters for the model to recognize.

3. Methodology and Challenges

In our attempt to extend the application of our character recognition model to the realm of medical prescriptions, we encountered unforeseen challenges that influenced the obtained results. The crux of the issue lay in the segmentation phase, where the prescription text is partitioned into individual characters for recognition.

Segmentation Challenges

The process involves various steps, such as grayscale conversion, thresholding, and contour detection, followed by character extraction and sorting. However, the variability and complexity of prescription handwriting posed difficulties, leading to imperfect segmentations. Notably, instances were observed where a character and a half, or even partial characters, were detected instead of complete ones. This phenomenon adversely affected the model's ability to accurately interpret the prescription text. For example, a character and a half instead of a single character were often segmented, which led to inaccuracies in recognition as shown in the following examples:

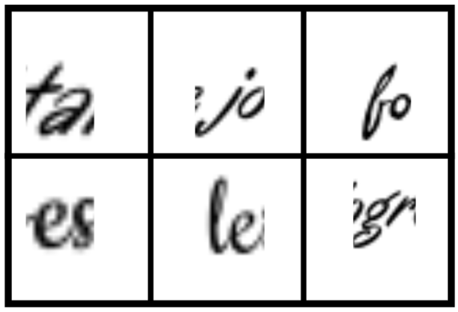

Fig. 7: Examples of incorrect segmentation

This missegmentation underscores the complexity of moving from controlled, individual character recognition to the more chaotic realm of full-text extraction. It highlights the need for sophisticated segmentation techniques that can more accurately separate and identify characters within free-form text.

Database Limitation

Compounding the segmentation challenges, the original database's limited diversity in handwriting styles hindered the model's adaptability to the intricate nature of medical prescriptions. Recognizing this constraint, we made a strategic decision to transition from our initial dataset to the "EMNIST" (Extended MNIST) open-source database.

4. Transition to EMNIST Database: A Strategic Shift

In response to the challenges encountered with the original dataset, we undertook a strategic shift to leverage the "EMNIST" (Extended MNIST) database. This transition aimed to address the limitations posed by the diversity of handwriting styles in medical prescriptions.

Introducing EMNIST Database

EMNIST, an extension of the well-known MNIST dataset, offers a diverse and expansive collection of handwritten characters. It encompasses a wide range of writing styles, reflecting the intricacies found in real-world documents. The dataset is organized into multiple splits, including ByClass, ByMerge, Balanced, Letters, and Digits, allowing us to tailor our focus to the specific demands of medical prescription recognition.

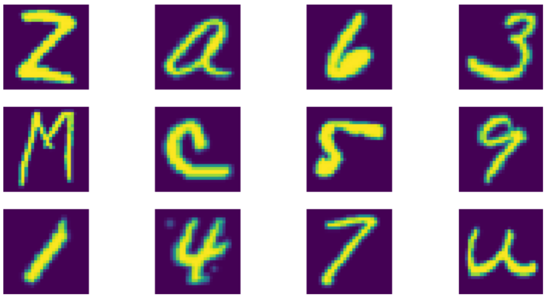

Fig. 8: Sample Characters from EMNIST Database

To provide a visual insight into the EMNIST (Extended Modified National Institute of Standards and Technology) database,this is a representative image showcasing a selection of characters. This image serves as a snapshot, offering a glimpse into the diverse set of characters present in the EMNIST dataset. It is important to note that the dataset comprises actual characters rather than pixel-level representations, allowing us to explore and understand the rich variety of handwritten symbols encapsulated within the EMNIST collection. This visual representation sets the stage for a more comprehensive exploration of the dataset's characteristics and our model's performance.

Training on EMNIST: Methodology

The training process involved adapting our existing character recognition model to the nuances of the EMNIST dataset. Preprocessing steps, including resizing and normalization, were adjusted to align with the characteristics of EMNIST images. Additionally, the model underwent retraining to capture the rich variety of writing styles present in the new dataset.

Results and Model Performance

The outcomes of training on the EMNIST dataset yielded promising results. The model exhibited improved adaptability to diverse handwriting styles, showcasing an impressive accuracy of 86% in recognizing characters typical of medical prescriptions.

Training Metrics: Loss and Accuracy

To provide a comprehensive view of the training process, we present the following graphs illustrating the model's loss and accuracy trends throughout the training epochs on the EMNIST dataset.

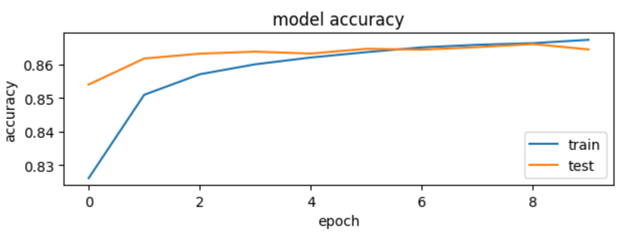

Fig. 9: Training and Validation Accuracy

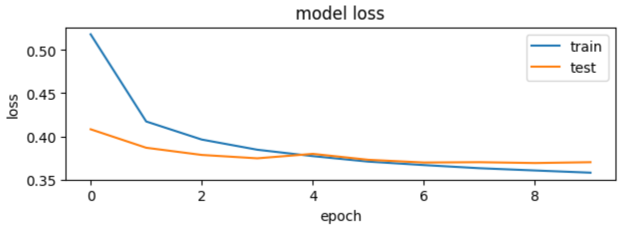

Fig. 10: Training and Validation Loss

These visual representations offer insights into the model's learning dynamics, emphasizing the effectiveness of the EMNIST dataset in refining our character recognition capabilities.

Image Segmentation: Ensuring Precision

Prior to character recognition, our pipeline incorporates a robust image segmentation process to delineate individual characters within medical prescriptions. The segmentation technique effectively identifies and extracts distinct characters, laying the foundation for accurate recognition.

Below is an illustrative example demonstrating the effectiveness of our image segmentation:

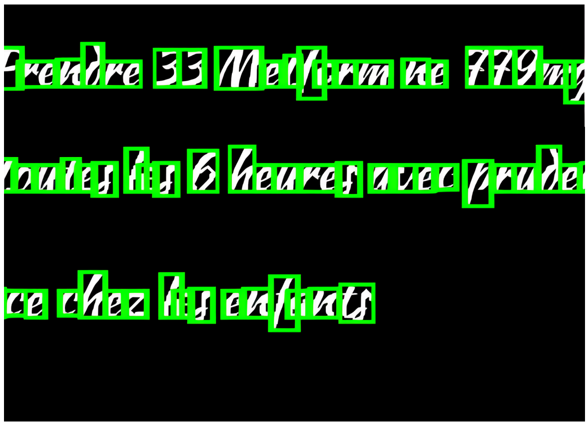

Fig. 11: Image showcasing the good segmentation of a medical prescription

The next figure provides a side-by-side comparison of the original text from a medical prescription with the text recognized by our character recognition model trained on the EMNIST dataset. The recognized text highlights the model's commendable ability to accurately capture the majority of the content, albeit with minor discrepancies. Notably, "Metformine 779mg" is recognized nearly accurately as "Metformine 779mg," and "toutes les 6 heures avec prudence chez les enfants" is correctly identified, showcasing the model's proficiency. However, the misrecognition of "t" as "d" in "chez" points to the ongoing challenges in character differentiation within our trained model.

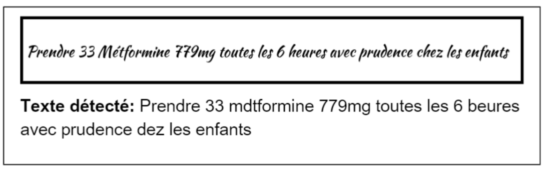

Fig. 12: Image showcasing the good segmentation of a medical prescription

IV. Discussion and Synthesis

Evolution of Approach

The evolution of our approach reflects a dynamic process aimed at enhancing the efficacy of our OCR system for medical prescriptions. Initially, we generated a diverse character dataset and trained a CNN model, achieving a commendable accuracy of 90%. However, when applied directly to medical prescriptions, the model exhibited limitations in segmentation and recognition.

Limitation of Character Dataset

The character dataset initially employed, while diverse, revealed limitations in capturing the intricacies of medical prescription text. The dataset's scope may not have adequately represented the wide array of characters, symbols, and variations encountered in real-world prescriptions. This gap in dataset diversity contributed to suboptimal performance when applied to the specialized domain of healthcare documentation.

Segmentation Challenges

Furthermore, the segmentation challenges were not solely confined to character confusion. The model struggled with accurately segmenting characters within the context of prescription text, where variations in layout, font, and formatting abound. Instances of characters appearing as "half" or "merged" posed substantial obstacles to the accurate extraction of meaningful information. These segmentation difficulties highlighted the need for a more robust and adaptable approach tailored to the intricacies of medical prescriptions.

Transition to EMNIST Dataset

Recognizing the limitations, our strategic shift to the EMNIST open-source dataset aimed at addressing both the database insufficiency and segmentation challenges. The EMNIST dataset, designed for comprehensive character recognition tasks, offered a broader and more representative set of characters. This transition significantly contributed to mitigating the challenges associated with dataset limitations and improving the segmentation accuracy within the medical prescription context.

V. Conclusion

This study embarked on a comprehensive exploration of Optical Character Recognition (OCR) systems tailored for medical prescriptions, aiming to address the intricate challenges associated with character segmentation and recognition within this specialized domain. The journey began with the development of a custom CNN model trained on a diverse character dataset. While exhibiting high accuracy in character recognition, the model faced inherent challenges in accurately segmenting and interpreting medical prescription text.

The identified challenges, including character confusion and segmentation complexities, prompted a strategic shift in our approach. Transitioning to the EMNIST open-source dataset proved instrumental in overcoming the limitations of the initial character dataset and improving segmentation accuracy. The EMNIST dataset, designed for comprehensive character recognition, provided a more representative set of characters and contributed to the overall efficacy of the OCR system.

The findings of this study underscore the nuanced nature of medical prescription text, requiring specialized approaches for accurate extraction of information. The successful integration of the EMNIST dataset highlights the importance of dataset choice in OCR model development, especially in domains with distinct linguistic and typographic characteristics.

Looking ahead, future research will delve into further refining segmentation algorithms, leveraging advanced machine learning techniques, and exploring the synergy between custom-trained models and pre-existing OCR tools. The goal is to develop a more robust OCR system capable of handling the diverse layouts, fonts, and variations encountered in medical prescriptions.

In conclusion, this study contributes valuable insights to the evolving field of OCR systems for healthcare documentation. The challenges encountered and solutions proposed pave the way for future advancements, emphasizing the need for domain-specific considerations in developing effective OCR solutions for medical contexts.

REFERENCES

Preetha S , Afrid I M , Karthik Hebbar P and Nishchay S Kd:Machine Learning for Handwriting Recognition

Surya Nath R S , Afseena S:Handwritten Character Recognition – A Review

Polaiah Bojja , Naga Sai Satya Teja Velpuri, Gautham Kumar Pandala,S D Lalitha Rao Sharma Polavarapu , Pamula Raja Kumari:Handwritten Text Recognition using Machine Learning Techniques in Application of NLP

Cohen, G., Afshar, S., Tapson, J., van Schaik, A. (2017). EMNIST: an extension of MNIST to handwritten letters.