Medical Prescription Recognition System

Indian Institute of Information Technology, Guwahati
Advisor: Dr. Ferdous Ahmed Barbhuiya

Krishna Sharma
Roll no.: 2101104

This disseration is submitted for the degree of
Bachelors of Technology in Computer Science Engineering

IIIT GUWAHATI April 2024

Table of Contents

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Krishna Sharma
April 2024

Acknowledgement

I would like to thank the Department of CSE for going ahead and allowing me to pursue a project aligning with the directions of my interest. I would like to thank my advisor, Dr. Ferdous Ahmed Barbhuiya, for continuously guiding me throughout the project and motivating me to keep moving.

Abstract

This report addresses the pervasive challenge in the medical field posed by deciphering doctors' illegible handwriting, which presents a significant obstacle for both patients and pharmacists, potentially resulting in medication errors. In response to this issue, the report proposes the development of a recognition system aimed at translating physicians' handwritten prescriptions into clear and understandable text. Through the utilization of advanced image pre-processing and word segmentation techniques, prescription images are meticulously prepared for model training. Deep learning methodologies, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory networks (LSTMs), are employed to develop a robust deduction model. The overarching goal of this project is to enhance the readability and accuracy of prescriptions, effectively addressing the challenges posed by physicians' handwriting within healthcare settings. By streamlining the prescription interpretation process, this recognition system endeavors to improve medication management practices, ultimately contributing to enhanced patient safety and healthcare efficiency.

1 Introduction

In the realm of healthcare, the ability to decipher doctors' illegible handwriting is crucial for accurate medication management. However, amidst the demanding schedules of medical professionals, where time is a precious commodity and consultations are numerous, the focus often shifts more towards delivering precise diagnoses rather than meticulously crafting prescriptions. Consequently, doctors frequently exhibit poor handwriting, posing significant challenges for patients and pharmacists in interpreting prescriptions and identifying medications and dosages accurately.

The implications of medication errors extend across a broad spectrum, ranging from minor discomfort to severe complications, highlighting the paramount importance of clear and legible prescriptions in healthcare. The inability to interpret prescriptions accurately due to poor handwriting can result in incorrect medication administration, adverse drug reactions, and even life-threatening situations. Thus, there arises an urgent need for effective solutions that enhance prescription legibility and accuracy, thereby minimizing the risk of medication errors and improving patient safety.

Recognizing the significance of this issue, there is an increasing acknowledgment of the necessity for a tool that streamlines prescription interpretation processes. The advent of deep learning techniques, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory Networks (LSTM), presents promising opportunities for developing robust solutions in this domain. These advanced algorithms offer the potential to automate the recognition and conversion of handwritten prescriptions into digital text, revolutionizing medication management practices.

The primary objective of this study is to delve into the development of a sophisticated tool capable of actively recognizing medical prescription images and converting them into digital text. At the heart of this endeavor lies the utilization of ResNet18 architecture, a powerful deep learning model renowned for its effectiveness in image analysis tasks. By harnessing the capabilities of ResNet18 alongside advanced deep learning techniques, such as CNNs, RNNs, and LSTMs, the application aims to extract individual words from prescription images and recognize them to facilitate accurate prescription interpretation.

Through this innovative approach, the study endeavors to streamline medication management processes, making it more convenient for individuals to adhere to their prescribed dosages accurately. By leveraging the power of artificial intelligence and image processing, the application seeks to enhance efficiency and accuracy in prescription interpretation, ultimately benefiting both healthcare professionals and patients alike.

Overall, this report aims to explore the intricacies of developing a cutting-edge tool to address the challenges posed by poor handwriting in medical prescriptions. By providing a reliable and efficient solution for prescription interpretation, the study endeavors to improve medication management practices, minimize the risk of medication errors, and enhance patient safety within healthcare settings.

2 Related Literature

2.1 Recognition of Doctors' Cursive Handwritten Medical Words by using Bidirectional LSTM and SRP Data Augmentation (2021)[1]

The paper suggests an online handwritten recognition system to identify doctors' handwriting and create a digital prescription using machine learning techniques . The study developed a primary "Handwritten Medical Term Corpus" dataset with 17,431 data samples comprising 480 words from 39 Bangladeshi doctors. On the preprocessed pictures, a new data augmentation technique called SRP is used to increase the number of data samples. Following this, a sequence of line data is extracted from both the original and augmented image data. Bidirectional LSTM is applied to the sequential line data derived from the augmented handwritten images to produce complete end-to-end recognition.

2.2 Handwriting Recognition for Medical Prescriptions using a CNN-Bi-LSTM Model (2021)[2]

It is difficult to decipher a doctor's handwriting on a prescription. In this paper, they used neural network techniques such as CNN and BI-LSTM for predicting doctor's handwriting from medical prescriptions . The CTC loss function is used for normalization. This model builds on the IAM dataset. Image acquisition and data augmentation are used for image preprocessing. Furthermore, it is passed as input to 7 convolution layers of a neural network. 32 training epochs were used by the training model, which took six hours to complete training and, on a graph, loss values are represented.

2.3 Medical Handwritten Prescription Recognition Using CRNN (2019)[3]

The approach established a Convolutional Recurrent Neural Network (CRNN) technology using Python that can interpret handwritten English prescriptions and translate them into digital text . For this, datasets with 66 different classes, including alphanumeric characters, punctuation, and spaces, were used. Since prescriptions generally contain two or three words, the training was carried out using short texts. Normal handwriting and prescriptions from doctors were used to train the model. This paper further stated that in order to enhance the results, more work is needed on input handling techniques.

2.4 Intelligent Tool For Malayalam Cursive Handwritten Character Recognition Using Artificial Neural Network And Hidden Markov Model (2017)[4]

The approach uses the Hidden Markov Model (HMM) to recognize cursive handwritten Malayalam characters . By employing a median filter, the algorithm used here helps to avoid errors caused by noise in the scanned image. Furthermore, Artificial Neural Network (ANN) aids in the acquisition of better classification and provides the best matching class for input. The samples used are of high quality in order to reduce the complexity of the recognition process. This method yields better results in terms of speed and accuracy. As a result, the combination of both English and Malayalam characters can be recognized as future work.

3 Proposed Solution

The project at hand encompasses two distinct yet interconnected components: word detection and recognition. The initial phase, word detection, revolves around the segmentation of individual words from the scanned prescription image. This crucial step is executed using ResNet18, a convolutional neural network renowned for its prowess in image recognition tasks. ResNet18 efficiently identifies and isolates each word, laying the foundation for subsequent processing.

Moving forward, the focus shifts to word recognition, the process of deciphering the actual meaning of each isolated word. This intricate task is accomplished through a sophisticated architecture comprising Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Connectionist Temporal Classification (CTC) layer. This layered approach harnesses the collective power of deep learning to accurately interpret the handwritten text, providing valuable insights into the prescribed medications.

By seamlessly integrating these two components, the project aims to streamline the prescription interpretation process.

3.1 Data Collection and Preprocessing

Data plays a pivotal role in the development and evaluation of our deep learning model. We have leveraged the IAM dataset, a widely recognized resource known for its comprehensive collection of handwritten text samples, to fuel our project.

- **Gray-scale conversion:** Gray scaling is the first step in digital image pre-processing that must be done. Here each pixel's value solely encodes the light's intensity information. There are just three colors used in grayscale images: black, white, and gray, which comes in a variety of tints.
- **Normalization:** Following grayscale conversion, the images undergo normalization—a critical process aimed at refining the data for seamless integration into the model. Normalization involves scaling down the pixel values from their original range of 0 to 255 to a standardized range of 0 to 1. This uniformity ensures consistency across the dataset, reducing potential discrepancies and facilitating effective model training.

3.2 Word Detection

In this part we try to extract the words from an image.The model operates by meticulously classifying each pixel within an image, discerning whether it belongs to the inner part of a word, the surrounding area of a word, or constitutes background. For pixels identified as part of the inner word, the model predicts an axis-aligned bounding box (AABB) encompassing the word. Given the potential for multiple AABBs to be predicted for the same word, a sophisticated clustering algorithm is employed to effectively group them.

The output maps of the model encode the AABBs and their respective classifications. These output maps include:

- Three segmentation maps utilizing one-hot encoding to designate:
  - The inner part of a word
  - The surrounding area of a word
  - Background pixels
- Four geometry maps that encode distances between each pixel and the edges of the AABB, specifying:
  - The distance to the top edge

– The distance to the bottom edge

– The distance to the left edge

– The distance to the right edge

ResNet18 serves as the feature extractor, following the widely recognized U-shape architecture commonly used in segmentation tasks. During the training process, the input images are resized to dimensions of 448×448 pixels. After traversing through the final layer of ResNet18, the feature maps undergo a downscaling process, resulting in dimensions of 14×14. Subsequent layers progressively upscale these maps, amalgamating intermediate results from ResNet18 and extracts the predicted bounding box. Ultimately, the output of the neural network yields all the predicted bounding boxes in the image.
The loss function comprises two integral components essential for effective training:

- Segmentation loss: To tackle the pixelwise classification problem inherent in segmentation tasks, cross-entropy loss is utilized to ascertain the accuracy of segmenting each pixel.
- Geometry loss: Rather than relying on sum-of-squared errors on the geometry, which may disproportionately weigh larger bounding boxes, the intersection over union (IOU) metric is adopted. This metric ensures a fair assessment of the overlap between predicted bounding boxes and ground truth boxes.

To address situations where multiple AABBs are predicted for the same word, a robust clustering algorithm named DBSCAN is deployed. This algorithm calculates AABB clusters based on the Jaccard distance between AABB pairs. Subsequently, the resulting AABB for each word cluster is meticulously computed, taking into account the median edge positions of its constituent members. This meticulous process ensures accurate localization of words within the image, laying a strong foundation for effective word recognition. The output of this network are fed to the predicting model.

3.3 Word Recognition

In this part we try predict the accurate text from the image. Our neural network (NN) architecture comprises convolutional NN (CNN) layers, recurrent NN (RNN) layers, and a final Connectionist Temporal Classification (CTC) layer, each playing a crucial role in predicting the word from the input image.

CNN Layers

- The input image undergoes processing through the CNN layers, which are adept at extracting pertinent features from the image.
- Each CNN layer consists of three operations: convolution, rectified linear unit (RELU) activation, and pooling.
- Through convolution, filters of varying sizes (5x5 in the initial layers and 3x3 in subsequent layers) are applied to the input, capturing spatial patterns.
- RELU activation introduces non-linearity, enhancing the network's ability to capture complex relationships within the image.
- Pooling layers summarize image regions, downsizing the input while preserving essential features.
- The output of the CNN layers is a feature map or sequence with dimensions of 32x256, where each element represents a feature extracted from the image.

RNN Layers

- Operating on the feature sequence derived from the CNN layers, the RNN propagates relevant information through the sequence.
- Leveraging Long Short-Term Memory (LSTM) cells, the RNN can effectively capture dependencies over longer distances and exhibits robust training characteristics.
- LSTM cells are designed to overcome the vanishing gradient problem commonly encountered in traditional RNNs. This problem arises when gradients become extremely small during backpropagation, hindering the learning process, especially for long sequences.
- It employs gating mechanisms, including input, forget, and output gates, to regulate the flow of information within the cell.
- The RNN output sequence is transformed into a matrix of dimensions 32x80.

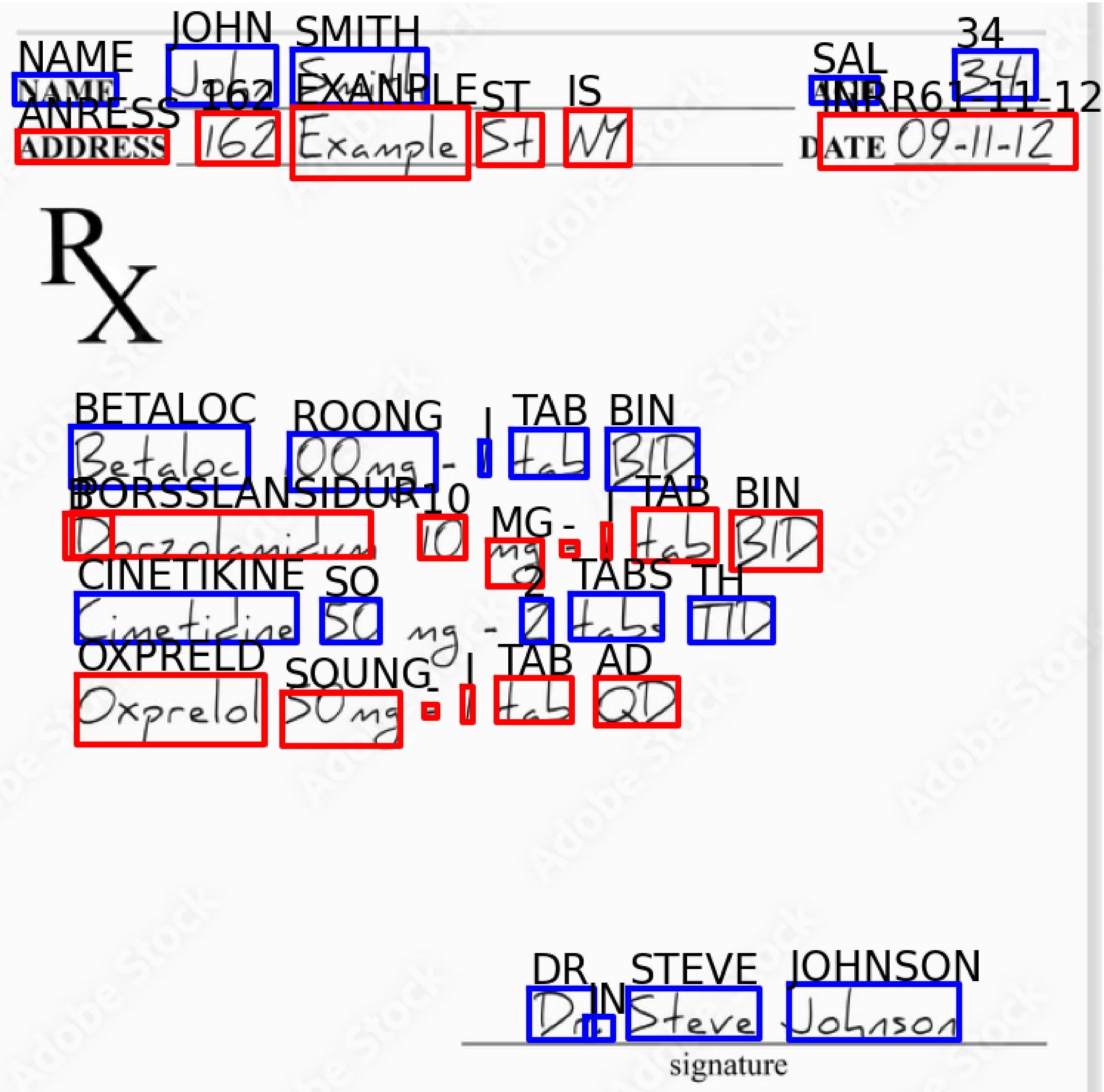
CTC Layer

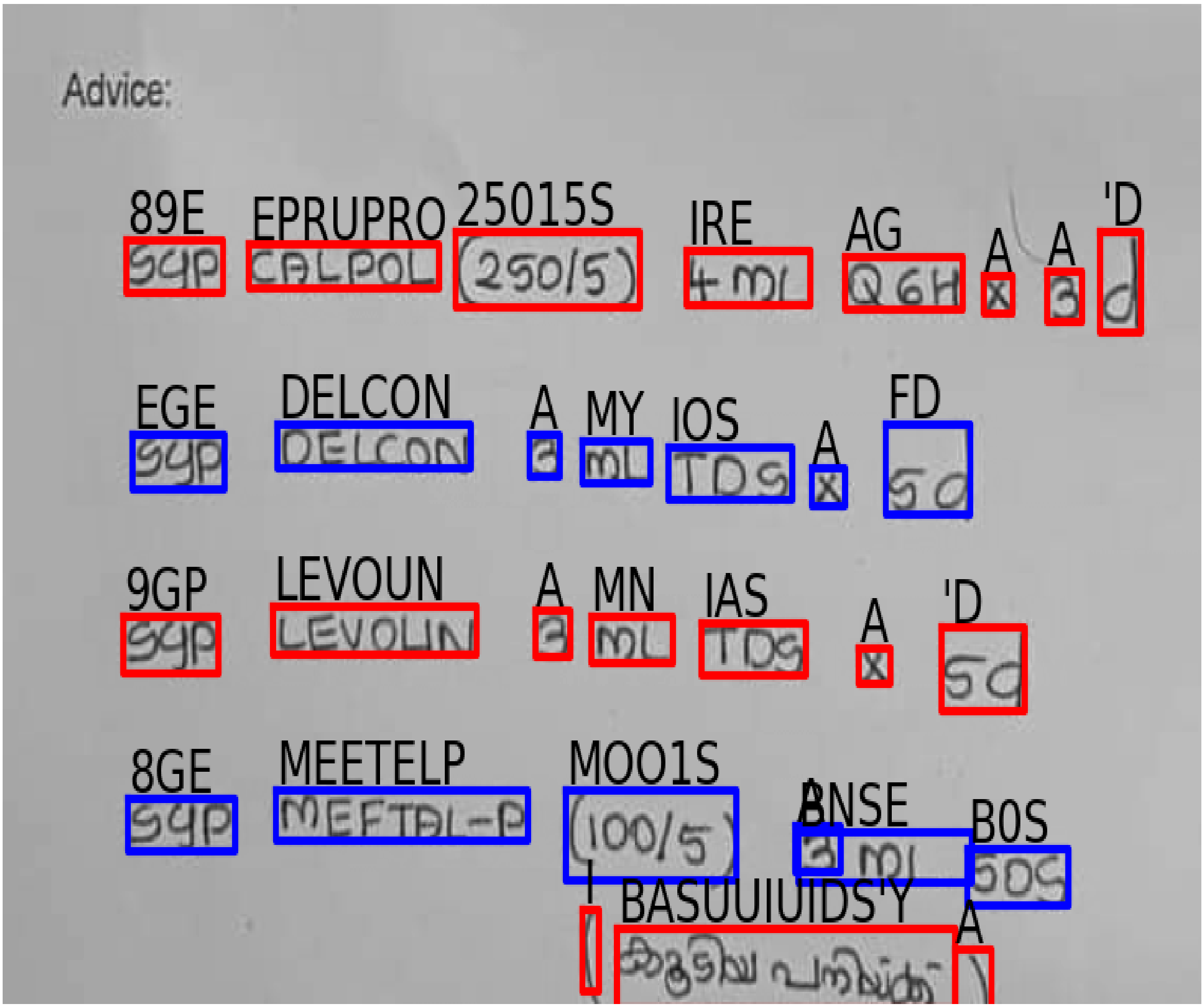
- During training, the CTC layer receives the output matrix from the RNN layers and the ground truth text, computing the loss value.
- In inference mode, the CTC layer decodes the output matrix into the final predicted text.
- CTC layer processes input sequences and calculates probabilities of target output sequences given input, allowing for variable-length inputs and outputs.
- Notably, both ground truth and predicted texts are constrained to a maximum length of 32 characters.

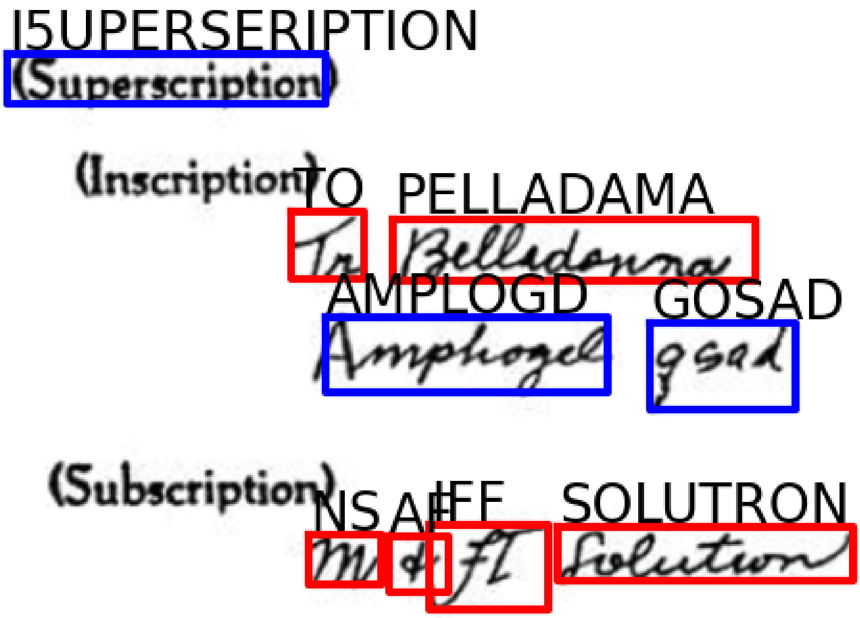
Each component of the NN architecture collaborates synergistically to process the input image, extract relevant features, capture temporal dependencies, and decode the output matrix into the predicted text, ultimately enabling accurate word prediction.

4 Results

By using the above mentioned technique we can predict the text with an average accuracy of nearly 57%. Some results are




Advice:
89E
EPRUPRO
25015S
IRE
AG
A
A
'D
EGE
DELCON
A
MY
IOS
A
FD
9GP
LEVOUN
A
MN
IAS
A
'D
8GE
MEETELP
MOO1S
BNSE
B0S
I
BASUUIUIDS'Y
A


I5UPERSERIPTION
(Superscription)
(Inscription)
TO
PELLADAMA
Belladonna
AMPLOGD
GOSAD
(Subscription)
NS
AF
IFF
SOLUTRON
Solution

5 Conclusion

In conclusion, the presented approach utilizes a combination of convolutional neural networks (CNN), recurrent neural networks (RNN), and connectionist temporal classification (CTC) to achieve word prediction in handwritten images. By leveraging the capabilities of ResNet18 for feature extraction, CNNs and RNNs for sequence modeling, and CTC for loss computation and decoding, the model demonstrates promising results in predicting words from scanned images of handwritten text.

Through the implementation of this approach, several key insights have been gained into the challenges and opportunities of handwritten word prediction. By successfully classifying pixels, predicting bounding boxes, and decoding sequences, the model showcases the potential of deep learning techniques in tackling complex image processing tasks.

Moving forward, there are several avenues for future exploration and improvement:

6 Future Work

- **Model Optimization:** Further optimization of the neural network architecture and hyperparameters could enhance the model's performance and efficiency.
- **Dataset Expansion:** Increasing the diversity and size of the training dataset, possibly by incorporating additional handwriting styles and languages, can improve the model's generalization capabilities.
- **Fine-tuning Techniques:** Implementing fine-tuning techniques such as transfer learning or domain adaptation could enable the model to adapt more effectively to specific handwriting styles or domains.
- **Integration with Applications:** Integrating the trained model into practical applications, such as handwriting recognition software or document processing systems, could facilitate real-world usability and impact.
- **User Interface Enhancement:** Improving the user interface and user experience of the system, including error handling and feedback mechanisms, can enhance usability and accessibility.

References

[1] S. Tabassum et al., ”Recognition of Doctors’ Cursive Handwritten Medical Words by using Bidirectional LSTM and SRP Data Augmentation,” 2021 IEEE Technology Engineering Management Conference - Europe (TEMSCON-EUR), Dubrovnik, Croatia, 2021, pp. 1-6, doi: 10.1109/TEMSCON-EUR52034.2021.9488622.

[2] T. Jain, R. Sharma and R. Malhotra, ”Handwriting Recognition for Medical Prescriptions using a CNN-Bi-LSTM Model,” 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-4, doi: 10.1109/I2CT51068.2021.9418153.

[3] R. Achkar, K. Ghayad, R. Haidar, S. Saleh and R. Al Hajj, ”Medical Handwritten Prescription Recognition Using CRNN,” 2019 International Conference on Computer, Information and Telecommunication Systems (CITS), Beijing, China, 2019, pp. 1-5, doi: 10.1109/CITS.2019.8862004.

[4] N. P. T. Kishna and S. Francis, ”Intelligent tool for Malayalam cursive handwritten character recognition using artificial neural network and Hidden Markov Model,” 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 2017, pp. 595-598, doi: 10.1109/ICICI.2017.8365201.