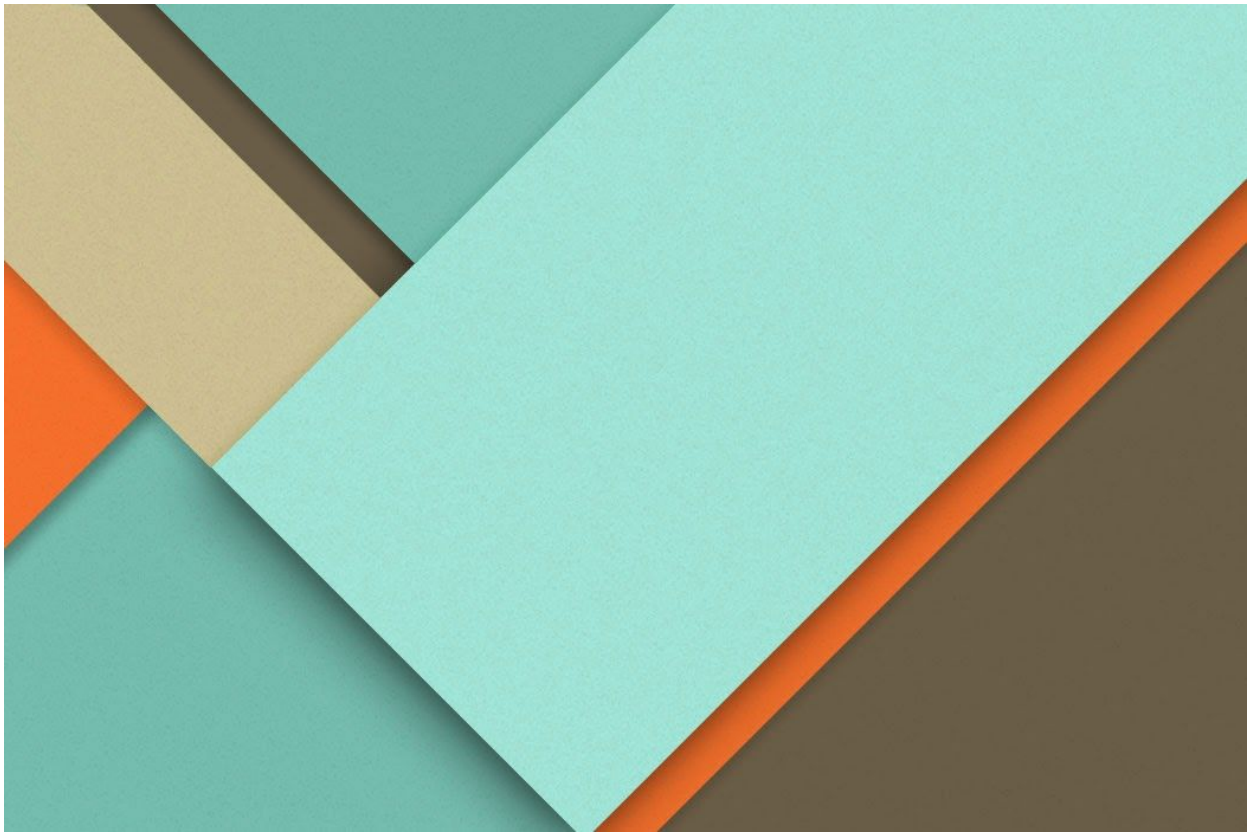


Notebook To Notepad

24.05.2020

Ishika Agarwal
18103034

Summary

There are various techniques to store the information available in handwritten text documents into a computer storage disk and then later reusing it. Among all those techniques, we have chosen Optical Character Recognition (OCR) as the main fundamental technique.

Our project focuses on digitizing printed texts so that they can be electronically searched, stored, edited. Handwritten notes, such as letters, school notes, diaries, meeting minutes, grocery lists, recipes etc can be scanned and converted into text, available for use in any digital platform.

This project is a part of CSN206 Engineering analysis and design and CSN 208 Database Management system.

Table Of Contents

Introduction

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. One of many such areas is the domain of Computer Vision. The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. We have developed a model which uses such techniques to make keyboarding simpler.

At the present time, keyboarding remains the most common way of inputting data into computers. This is probably the most time consuming and labour intensive operation. We have chosen Optical Character Recognition (OCR) as the main fundamental technique to recognize handwritten texts. It is the technique used to convert scanned images which are handwritten to electronic form.

Our project focuses on digitizing printed texts so that they can be electronically searched, stored, edited. Handwritten notes, such as letters, school notes, diaries, meeting minutes, grocery lists, recipes etc can be scanned and converted into text, available for use in any digital platform.

A document scanned using a scanner and converted to an image will be input to the “Notebook to Notepad” system which recognizes the characters in the scanned documents and converts them into ASCII data. The quality of the scanned document depends upon the scanner. So, a scanner with high speed and colour quality is desirable.

An IAM database is used at the backend for recognition. The recognizing process includes use of several complex algorithms and previously loaded templates and dictionary.

Technical Requirements

The various technical skills were required to complete the project successfully.

NumPy

NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed.

Tensorflow

TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of TensorFlow. It can run on single CPU systems, GPUs as well as mobile devices and large scale distributed systems of hundreds of machines.

OpenCV

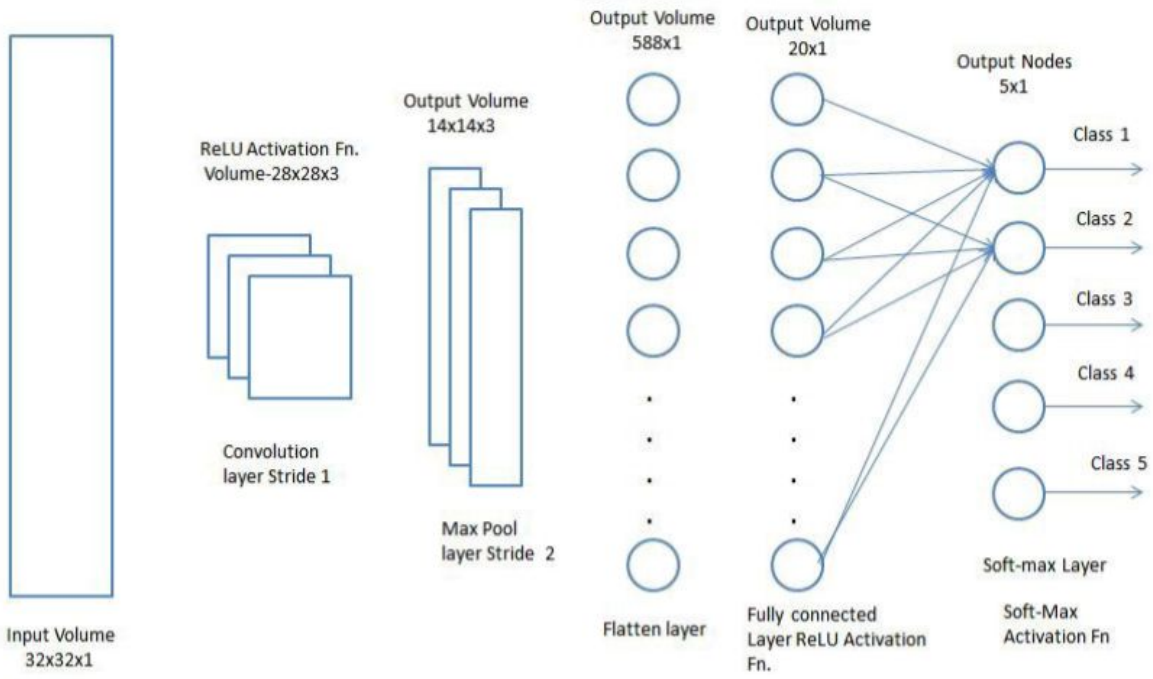
OpenCV is a huge open-source library designed to solve computer vision problems, machine learning, and image processing. OpenCV supports a wide variety of programming languages like Python, C++, Java, etc. It can process images and videos to identify objects, faces, or even the handwriting of a human.

Convolution Neural Networks

The advancements in Computer Vision with Deep Learning has been constructed and perfected with time over one particular algorithm called a Convolutional Neural Network.

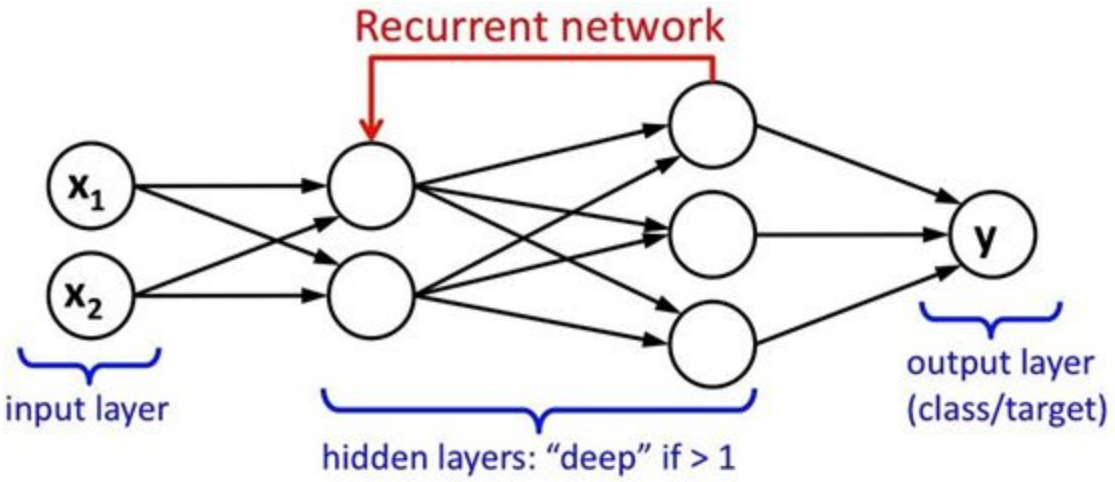
It takes in an input image, assign importance (learnable weights and biases)

to various objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms.The general structure of the network is shown below:



Recurrent Neural Networks

Recurrent Neural Network RNN) are a type of Neural Network where the **output from the previous step is fed as input to the current step**. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is **Hidden state**, which remembers some information about a sequence.



Connectionist Temporal Classification

Connectionist temporal classification (CTC) is a type of neural network output and associated scoring function, for training recurrent neural networks (RNNs) such as LSTM networks to tackle sequence problems where the timing is variable. It can be used for tasks like on-line handwriting recognition[1] or recognizing phonemes in speech audio.

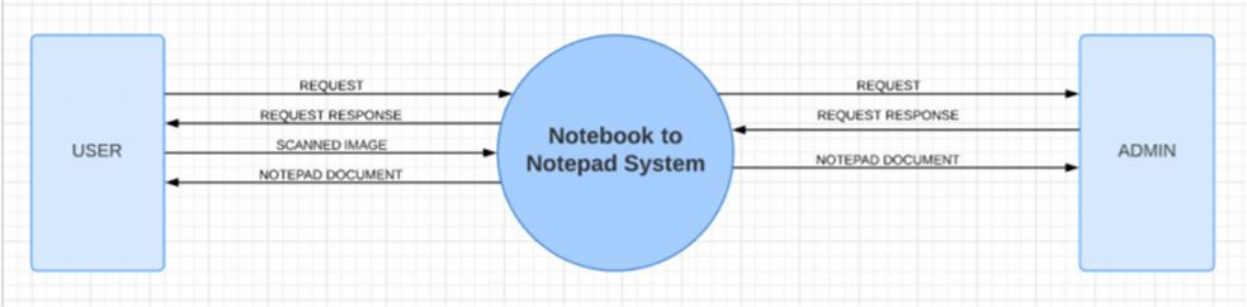
The input is a sequence of observations, and the outputs are a sequence of labels, which can include blank outputs.

System Design

We designed the architecture, modules, interfaces, and data for the system to satisfy specified requirements. The system was designed using:

1. **Data Flow Diagrams (DFDs)** consisting of three levels, a zero-level data flow diagram describing the relation between the information system as a whole and the entities involved in the project; first-level data flow describing the main components of the information system, their interactions with each other, entities and the databases involved; and second-level data flow diagram describing the further division of complex systems into simpler systems working synchronously to process the inputs and give desired outputs.
2. **Decision Tree** describing the condition-action relation upon which the system works, thus helping to understand which conditions lead to the system producing the result in a given way, and how to change those conditions to get the desired result.

Data Flow Diagram - Level 0



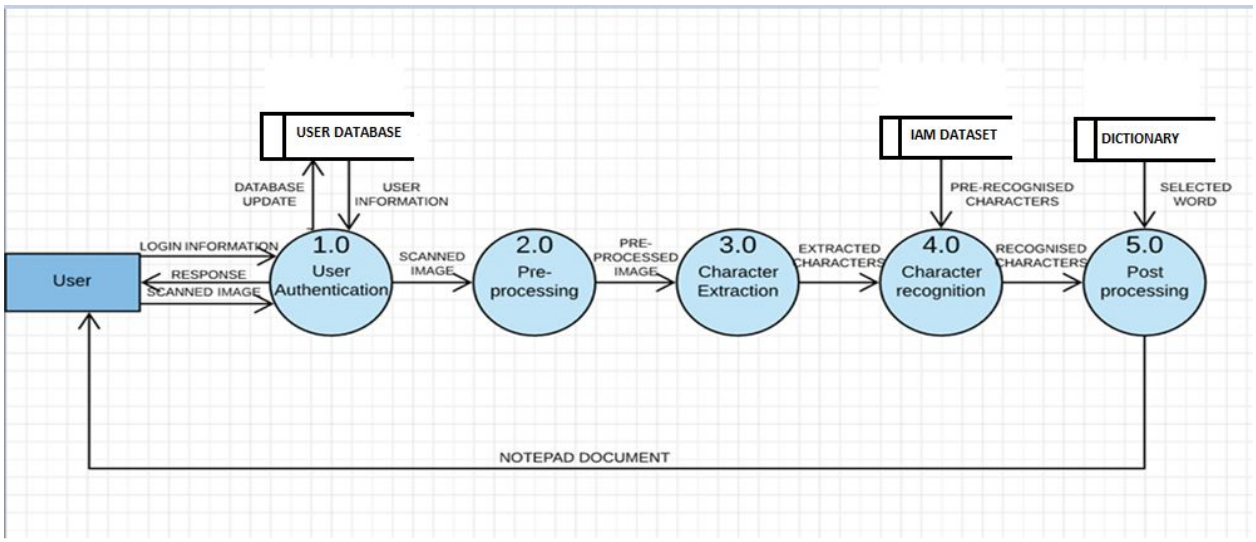
The entities in level 0 are:

1. **User:** The end user who will interact with the front end of the web-based information system. The user sends a request to the system and gets a response. Based on this response the user sends an input image and gets a corresponding notepad document in return.
2. **Admin:** The admin receives the user's request and sends an appropriate response. The notepad document generated by the system is also returned to the admin.

Processes:

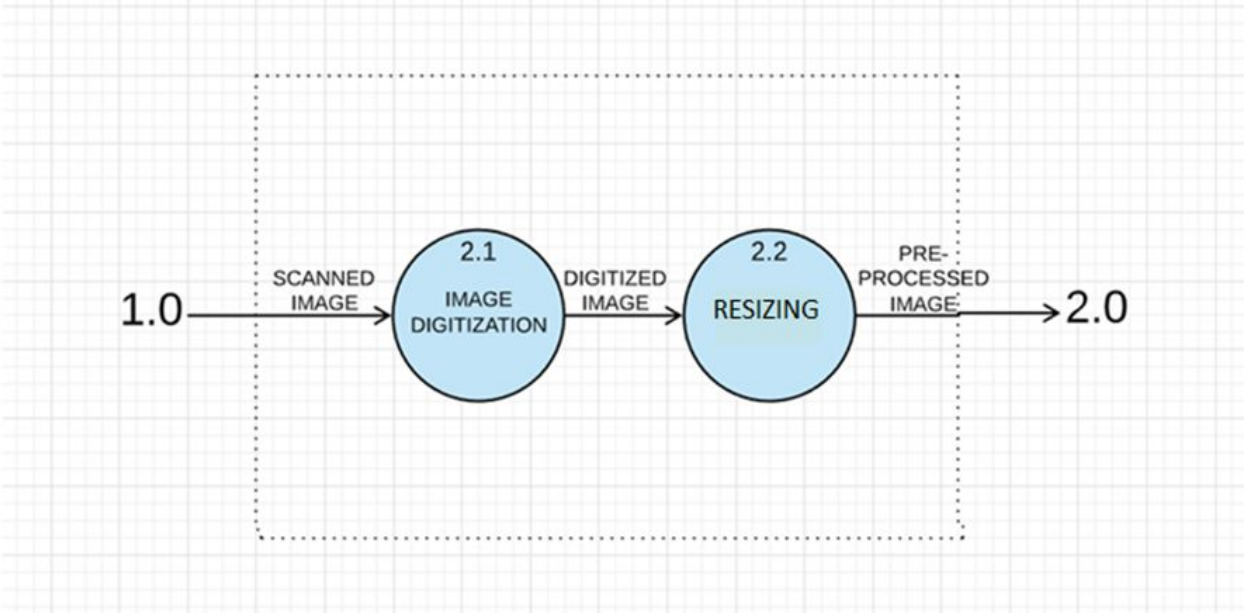
1. **Notebook To Notepad System:** It has 2 parts. One is the authentication system, and the other is a deep neural network which converts an image into a notepad file.

Data Flow Diagram - Level 1



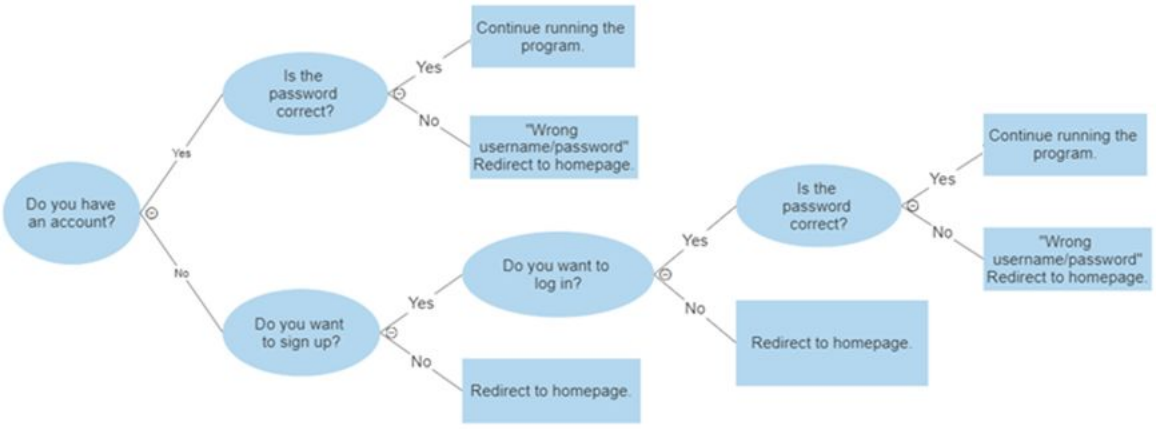
1. **User Authentication:** It authenticates the login information of the user and updates the User database.
2. **Pre-processing:** It pre-processes the scanned image. It removes noise and reduces size of the image.
3. **Character extraction:** The pre-processed image serves as the input to this and each single character in the image is found out.
4. **Character recognition:** The image from the extraction stage is correlated with all the templates which are preloaded into the system. Once the correlation is completed, the template with the maximum correlated value is declared as the character present in the image.
5. **Post processing:** If there are some unrecognised characters found, those characters are given their meaning in the post-processing stage. Word beam search is applied to find the best fitting word.

Data Flow Diagram - Level 2



1. **Image digitization:** Converts the scanned image to digitised form using CV2.
2. **Resizing:** Resizing is necessary as all the images in the dataset, and even the image the user would input are of different sizes. We have resized all the images to 800x64 pixels. The excess space is filled with 1s, essentially making the background white.

Decision Tree



First, the user is asked whether he already has an account or not. If he does, he can enter his login credentials. If they are correct, the program continues running and he is asked to upload an image. If not, he is redirected to the homepage.

If the user doesn't have an account, he can sign-up. After signing-up he is again redirected back to the login page, from where the same process follows.

Data Dictionary - Processes

Level 0

1. **Process Name:** Notebook To Notepad System

   Description: It takes a scanned image as the input, converts the text of this image to digitized form, stores it as a notepad file, and then gives the notepad document as the output.

   Inbound Data Flow: User request, scanned image, request response

   Outbound Data Flow: Request response, Notepad document

Level 1

2. **Process Name:** User Authentication (1.0)

   Description: It authenticates the login information of the user and updates the User database.

   Inbound Data Flow: Login information, scanned image, user information

   Outbound Data Flow: Response, Database update, Scanned image

3. **Process Name:** Pre-processing (2.0)

   Description: It pre-processes the scanned image. It removes noise and reduces size of the image.

   Inbound Data Flow: Scanned image

   Outbound Data Flow: Pre-processed image

4. **Process Name:** Character Extraction (3.0)

   Description: The pre-processed image serves as the input to this and each single character in the image is found out.

   Inbound Data Flow: Pre-processed image

Outbound Data Flow: Extracted characters

5. **Process Name:** Character Recognition (4.0)

   Description: The image from the extraction stage is correlated with all the templates which are preloaded into the system. Once the correlation is completed, the template with the maximum correlated value is declared as the character present in the image.

   Inbound Data Flow: Extracted characters, pre-recognised characters

   Outbound Data Flow: Recognised characters

6. **Process Name:** Post-processing (5.0)

   Description: If there are some unrecognised characters found, those characters are given their meaning in the post-processing stage. Word beam search is applied to find the best fitting word.

   Inbound Data Flow: Recognised characters, selected word

   Outbound Data Flow: Notepad document

Level 2

7. **Process Name:** Image Digitisation (2.1)

   Description: Converts the scanned image to digitised form.

   Inbound Data Flow: Scanned image

   Outbound Data Flow: Digitised image

8. **Process Name:** Noise Removal (2.2)

   Description: It removes unwanted noise from the image, which would lead to better learning by the neural network.

   Inbound Data Flow: Digitized image

   Outbound Data Flow: Pre-processed image

Data Dictionary - Data Flows

1. **Data Flow Name:** Login information

Description: It contains the username and password of that user.

From Process: NA

To Process: 1.0 User Authentication

2. **Data Flow Name:** Response

Description: It tells whether the user has been granted access or not.

From Process: 1.0 User Authentication

To Process: NA

3. **Data Flow Name:** Database Update

Description: It updates the User Database whenever a new user creates an account.

From Process: 1.0 User Authentication

To Process: NA

4. **Data Flow Name:** User Information

Description: It contains the actual password and details of said user. Its values is compared with Login Information to find out Response.

From Process: NA

To Process: 1.0 User Authentication

5. **Data Flow Name:** Scanned Image

Description: It is the scanned image input by the user whose digitised form will be returned.

From Process: 1.0 User Authentication

To Process: 2.0 Pre-processing

6. **Data Flow Name:** Digitised Image

Description: It converts the scanned imaged to digitised form.

From Process: 2.1 Image Digitisation

To Process: 2.2 Noise Removal

7. **Data Flow Name:** Pre-processed Image

Description: It is the pre-processed image, which has reduced size and removed noise.

From Process: 2.0 Pre-processing

To Process: 3.0 Character Extraction

8. **Data Flow Name:** Extracted Characters

Description: It is a list of all the characters that have been found in that image.

From Process: 3.0 Character Extraction

To Process: 4.0 Character Recognition

9. **Data Flow Name:** Pre-recognised Characters

Description: These characters are from the IAM Dataset and were used to train the Neural Network model.

From Process: NA

To Process: 4.0 Character Recognition

10. **Data Flow Name:** Recognised Characters

Description: These are the recognised characters of the image, obtained after comparing the pixel values of the image to the pre-recognised characters.

From Process: 4.0 Character Recognition

To Process: 5.0 Post-processing

11. **Data Flow Name:** Selected Word

Description: This is the word selected from the dictionary after applying word beam search.

From Process: NA

To Process: 5.0 Post-processing

12. **Data Flow Name:** Notepad Document

Description: This is the final output which is returned to the user.

From Process: 5.0 Post-processing

To Process: NA

Data Dictionary - Data Stores

1. **Data Store Name:** User Database

Description: It contains information about the user, including his name, username and password.

Inbound Data Flow: Database Update

Outbound Data Flow: User Information

Volume: Growing

Access: Can be accessed only by the admin

Data Description:

- first_name varchar(100)
- last_name varchar(100)
- username varchar(100)
- password varchar(100)

2. **Data Store Name:** IAM Dataset

Description: It contains a set of 115320 words, contributed by 657 writers. It has 1539 pages of scanned text, 5685 sentences and 13353 text lines. A mapping of these images with their corresponding text was created to train the Neural Network.

Inbound Data Flow: NA

Outbound Data Flow: Pre-recognised Characters

Volume: Static

Access: Read-only access while training the model. No access after that.

Conclusion

The project was able to complete its desired objectives which were to design and develop an information system that provides a way by which a person can convert handwritten text, cursive or capital to digitized form, eliminating the need for manually inputting data. This would save a lot of time and energy at the user's end, because keyboarding is a laborious task. Also, as each user has their own account, they can access all of their stored documents in one place, saving a lot of hassle.

We are trying to increase the accuracy of our model. The same can be done by adding more CNN and RNN layers.

References

The research in this area has been ongoing for over half a century and the outcomes have been astounding with successful recognition rates for printed characters exceeding 99%, with significant improvements in performance for handwritten cursive character recognition where recognition rates have exceeded the 90% mark.

The following are few projects we came across, which inspired us to take up this problem statement:

1. **Handwriting Optical Character Reader**

https://github.com/Breta01/handwriting-ocr

This project tries to create software for recognition of a handwritten text from photos (also for Czech language). It uses computer vision and machine learning. The initial input is a photo of a page with some text.

2. **Simple Handwriting Text Recognition System**

https://github.com/githubharald/SimpleHTR

This system is implemented with TensorFlow (TF) and trained on the IAM off-line HTR dataset. This Neural Network (NN) model recognizes the text contained in the images of segmented words.

3. **Medical Prescription OCR**

https://github.com/Aniket025/Medical-Prescription-OCR

This system allows a doctor to write his prescriptions the conventional way (i.e., using their pen and paper). From the scanned version of the prescription, a handwritten character recognition will be followed to capture the data (name of the patient, symptoms, findings, prescription of medicine, tests, advice, etc.) written by the doctor.