Extraction of Cursive Writing Character

J. Ajeetha Thasneem *1, D. Manibharathi *2

*1 PG Scholar-Communication Systems,Department of ECE,Government College of Engineering,Salem-11.
*2Assistant Professor,Department of ECE,Department of ECE,Government College of Engineering,Salem-11.

ABSTRACT
This Paper represents the work related to the extraction of cursive English handwriting. Character recognition(HCR) of handwritten cursive English script is a very challenging task. In handwriting, the characters in a word are connected. There should be an extreme amount of variations and overlapping because everyone has their unique writing style, characters are elongated, it may be Slanted, Stylized, Tiny, Gigantic, So Digitizing handwriting is extremely challenging and is still far from solved but Machine learning is helping us improve our handwriting recognition accuracy. In this project extracting the cursive handwritten words by using python and Open-CV, kernel filters are used for distorted the text from the image, and Using Git-hub for writing the code, then execution Google co laboratory is used, While developing this model it may use for School and college students to digitizing their handwritten notes
Keywords : Open CV,HCR,Google Co-lab,Git-hub

I. INTRODUCTION
Handwritten character recognition (HWR) is an ongoing field of research encompassing artificial intelligence, computer vision, and pattern recognition. An algorithm that performs handwriting recognition will acquire and find the characteristics from photos and convert them to a machine-readable form. The human sensory system is primarily concerned whenever individuals are reading Handwriting characters, letters, words, or digits. It seems simple whenever one is reading handwriting, but it is not as easy everyone understand. Humans can make sense of what they see based on what their brains have been taught, although everything is done unconsciously. A human will be appreciate how difficult it is to solve handwriting. The challenge of visual pattern recognition is only apparent to develop a computer system to read handwriting. The artificial neural networks approach is considered the best way to develop systems for recognizing handwriting. Neural networks facilitate to simulate however the human brain works when reading handwriting in a more simplified form. It allows machines to match and exceed human capabilities at reading handwriting. Humans have different handwriting styles, some of which are difficult to read.

1.1 Traditional techniques:
a.   Character extraction
In feature extraction stage every character is represented as a feature vector, that becomes its identity. The main goal of feature extraction is to extract a group of features, which maximizes the popularity rate with the quantity amount of components

b.   Character recognition
After the extraction of individual characters happens, a recognition engine is used to employed the corresponding computer character.

c.   Feature extraction
Feature extraction works in a similar fashion to neural network recognizer. However, programmers should manually determine the properties they feel are important. This approach offers the recognizer more control over the properties used in identification. Yet any system using this approach needs considerably additional requires substantially more development time than a neural network because the properties are not learned automatically.

d.   Modern techniques
Where ancient techniques focus on segmenting individual characters for recognition, modern techniques target recognizing all the characters in a segmented line of text. Significantly they focus on machine learning techniques that are able to learn visual features, avoiding the limiting feature engineering previously used.

II. METHODOLOGY
Programming will offer to the unimaginable things that are once not Impractical. As the world is developing so is the innovations and strategies to do new things. One such factor is machine learning and its large choices like data analytics, predictions and

many other things. One of the fascinating things that we are going to discuss is the handwritten text recognition from the image.

There are various applications of extracting and recognizing handwritten text such as product indexing at a super mart, making digital copies of handwritten notes and help in developing an AI that can write like a human. This project has planned to make this model convert handwritten notes to digital copies that can be utilized in schools and colleges. Then going to use python and Open-CV for our task.

The steps included in extracting words from the image are:
1.Image Prepossessing where we make all images the same size and converting them to grayscale
2.Applying kernel/filter to the image and using threshold from Open-CV
3.Finding contours and making rectangles for appropriate sizes
4.Save the cropped images to make a data set
Using module from python which is OPEN CV which provides basic image processing features such as converting images, extracting digital OCR and editing images.
Flowchart of proposed system as below

First importing our image in Co laboratory then the image preprocessed and converted to Gray scale image for apply threshold we are using kernel filter, Then Finding contour function from Open CV is used to find the contours from the text and gives the cropped image Using OpenCV for most of our task. What understand from OCR extraction is that it uses a binary kind of system where all the black letters are separated from the white background or vice versa. That's why it is very important to rigorously preprocess our images. Below is the easy program that used to extract words from the image.

There are some easy imports, cv2 for OpenCV, numpy to handle image knowledge as an array as we know the image can be easily written in an array where each data point in the image array provides color data at every element, math to perform mathematical operations on image data, OS to handle files from the file system, matplotlib to show the image.Define two directories, one from where we will fetch all our images and one to save all the cropped words from the images. Will be looking at the code that will deal with one image at a time but can easily make a for loop to apply it to all the images.Image_dir is for operations to be performed on single image as a demo.

2.1 Image preprocessing
Prepare image is a function that will change the height of the image and convert it into a grayscale image. While Converting the image into grayscale then only the threshold function done smoothly,First we use a kernel filter that is a filter which enhance our black and white portions on the image.Then read an image from the directory using imread function. Using the prepare_image function to convert our image into grayscale and get appropriate height in pixels. Use the kernel in a filter to make our image ready to apply a threshold.Image preprocessing

2.2 Kernel Filter: In image processing kernel is a two dimensional matrix of numbers,kernel is also known as convolution matrix or masks.The general purpose of this kernel is filtering and moving the masks from point to point image.Commonly filters are perform certain operations in the image
For Example: The following matrix is a image which has 6*6 and the value in the matrix is intensity of the pixel

Image(6*6)n*n   output:

-5	-4	0	8
10	2	2	3
0	-2	-4	-7
-3	-2	-3	-16

In this example the image is convoluted with the filter,convolution is nothing but it is misnomer and study the process of the image,6*6 matrix is multiplied with filter and the size of output should be considered as

(n-f+1)*(n-f+1) = (6-3+1)*(6-3+1) = 4*4

For 1st matrix the calculation is

3*1+0*0+1*(-1)+1*1+5*0+8*(-1)+2*1+7*0+2*(-2)=3+(-1)+1+(-8)+2+(-2)

=3-1+1-8+2-2

=3-8 =-5

As this follow the same steps for all matrix to get the output matrix which is smoothed of the raw image

Kernel filter provides the low and high pass filtering,It smoothing and Sharpening the image,then it removes any pixels that are darker than a certain fraction of the darkest neighboring pixel,This filter extracts the feature of the image,While considering the scale of the image some features range widely large in size

2.3 Finding contours and drawing rectangles around words

Once image makes use of threshold,then the image is ready to find contours in that, identifying the ones white blobs as a phrase. First, we use the findcontour characteristic from OpenCV to find the contours from the image. We initialize an empty listing where all of the recognized words and its corresponding contour will be identified. We undergo every contour this is identified and check for the contour region. We don't need all of the contours from the image, we simply want this is great sufficient to make a contribution for a word that's why we neglect about those contours with much less than the place of 350 pixels. If the contour location is great, we make a rectangle round it and identify that as a phrase image. We capture this individual phrase image and its corresponding rectangle inside the listing and we sort it. Now from this listing, we are able to extract our words using the imwrite function from OpenCV and we are marking those rectangles at the pre-processed grayscale image to

expose where all the phrases are diagnosed. in the end, we put all this data into a list simply in case we need any records about any image.

III. RESULTS AND DISCUSSION

While discussing results below image is imported in the directory as input image,Once the image imported convert the image to appropriate height


Once the image was imported, converting the image into grayscale so that the filter and threshold applications can be done smoothly.


Start by creating a kernel that is a filter that we want to apply on our image that will enhance our black and white portions on the image. Read an image from the directory using the imread function. Use the prepare image function to convert our image into grayscale and get appropriate height in pixels.

The kernel in a filter to make our image ready to apply a threshold.

It uses threshold, ready to identify contours in the image identifying those white blobs as a word.
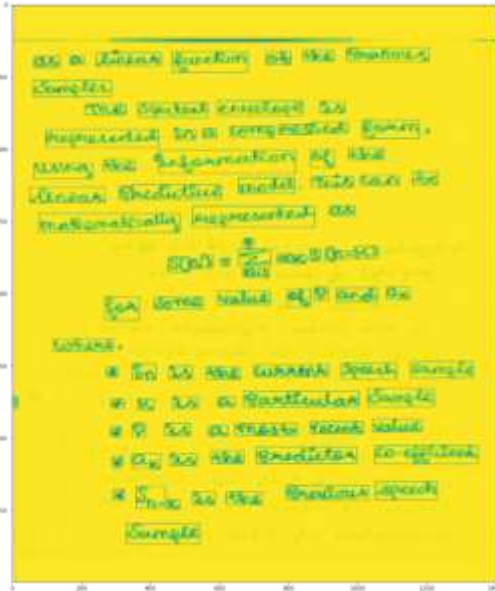
Once the contours are finder area more than specified range and respective rectangles have been drawn on the image.

The image which rectangles extracted and saved as a dataset in google drive
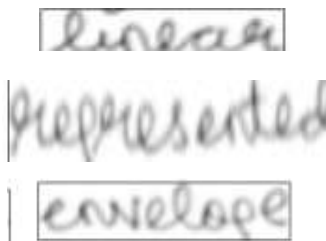
IV. CONCLUSION

By this model we can extract the words from the written handwriting document,Then using the extracted words as dataset, by using that we can predict the word which was written in document

V. FUTURE WORK

In Future can use these extracted words in any way we like, Here going use them to train a machine learning model to predict handwritten words from an image. By the help of these model, then going recognize the handwriting in medical prescription

There are various applications of extracting and recognizing handwritten text such as product indexing at a super mart, making digital copies of handwritten notes and help in developing an AI that can write like a human. It may also possible to convert handwritten notes to digital copies that can be utilized in schools and colleges.

REFERENCES
[1]. Aiquan Yuan, Gang Bai, Po Yang, Yanni Guo, Xinting Zhao College of Information Technical ScienceHandwritten English Word Recognition based on Convolutional Neural Networks,2012 International Conference on Frontiers in Handwriting Recognition
[2]. Zhang, G. Bai, and C. Xuan. A new methodology of online handwritten english character segmentation based on rules. In Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on, volume 1, pages 670–674. IEEE, 2010.
[3]. Abdulllah, M., Agal, A., Alharthi, M., & Alrashidi, M. (2018). Retracted: Arabic handwriting recognition using neural network classifier.
[4]. Journal of Fundamental and Applied Sciences, 10(4S), 265-270. [2] Abe, S. (2010). Support Vector Machines for Pattern Classification. Berlin, Germany: Springer Science & Business Media. Aggarwal, C. C. (2018).
[5]. Neural Networks and Deep Learning: A Textbook. Basingstoke, England: Springer. Balas, V. E., Roy, S. S., Sharma, D., & Samui, P. (2019).
[6]. Handbook of Deep Learning Applications. Basingstoke, England: Springer. Boukharouba, A., & Bennia, A. (2017).
[7]. Novel feature extraction technique for the recognition of handwritten digits. Applied Computing and Informatics, 13(1), 19-26. doi:10.1016/j.aci.2015.05.001
[8]. A.L. Bianne-Bernard, F. Menasri, R.A.H. Mohamad, C. Mokbel, C. Kermorvant, and L. Likforman-Sulem. Dynamic and contextual information in hmm model- ing for handwritten word recognition. Pattern Anal- ysis and Machine Intelligence, IEEE Transactions on, 33(10):2066–2080, 2011.
[9]. Zhou, Y. Lu, and C. Tan. Bangla/english script identification based on analysis of connected component profiles. Document Analysis Systems VII, pages 243– 254, 2006.
[10]. R.Yoganapriya, Prof.P.Indra., (2020). "DETECTION OF MASSES IN DIGITAL MAMMOGRAM USING NONLINEAR FILTERS", International Research Journal of Engineering and Technology (IRJET) Vol 07, Issue 12.

[11]. P.Indra (2020), "Multilevel Tetrolet transform based breast cancer classifier and diagnosis system for healthcare applications", in the proceedings of Journal of Ambient Intelligence and Humanized Computing ISSN 1868-5137, J Ambient Intell Human Comput.

[12]. P.Indra, (2015), "Survey on Mammogram Image Classification CAD system" in the proceedings of Global Journal of Engineering and Applied Sciences (GJEAS)

[13]. P.Indra, (2014), "Classification of Micro Calcification in Digital Mammogram using Tetrolet Transform", in the Proceedings of the International Journal of Innovative Research & Advanced Studies (IJIRS), Vol 3, Issue 8, pages:398-410, ISSN 2319-9728 on August 2014.

[14]. P.Indra, (2014), "Mammogram Classification based on conventional Harr Transform", in the Proceedings of the International Journal of Multidisciplinary Research and Advances in Engineering (IJMRAE), Vol 6 No IV, pages: 1-8, ISSN 0975-7074 on October 2014.