

**MIT- WORLD PEACE UNIVERSITY**

**PUNE, MAHARASHTRA**

**SCHOOL OF COMPUTER SCIENCE**

**2022-2023**

**Handwritten Devanagari Character Recognition**

**BY**

**Soham Kulkarni, Varun Kulkarni, Vachana Sholapur,**

**Rutuja Toravi, Prashant Donarkar**

**Under Guidance**

**Dr. C. H. Patil**

**Handwritten Devanagari Character Recognition using CNN**

**Soham Kulkarni, Varun Kulkarni, Vachana Sholapur, Rutuja Toravi, Prashant Donarkar**

*MIT World Peace University, Kothrud, Pune, Maharashtra 411057*

*Department of Computer Science and Technology*

**Abstract:**

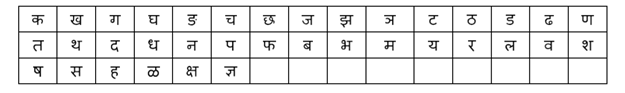
We present the implementation of Devanagari handwritten character recognition using CNN. Hand written Devanagari character recognition is gaining more importance due to its major contribution in automation system. Devanagari script is one of various languages script in India. It consists of 12 vowels and 36 consonants. Here we implemented the CNN model to recognize the characters. The character recognition consists of mainly five steps: pre-processing, segmentation, feature extraction, prediction, post-processing. The model will use convolutional neural network to train the model and recognize and predict the accuracy of recognition.

**Keywords:**

Devanagari Character Recognition, Convolutional neural network, Dataset.

**Abstract:**

Marathi is an Indo-Aryan language and also the official of Maharashtra. Marathi is usually written in balbodh version of Devanagari script. Recognizing these handwritten Devanagari characters is important because of its application in various fields such as bank cheque automation, postal automation, form processing, historical document preservation, etc. In India Devanagari script is most popular script used for the most spoken languages such as Sanskrit, Marathi and Hindi. Image acquisition , preprocesing, segmentation, feature extraction, classification, etc. are the major steps. Fig 1:- Sample Script.



**Convolutional Neural Network:**

CNN is a deep learning algorithm that is used to recognize and classify images. It is a type of deep neural network that requires very little preprocessing. It inputs the image in small chunks rather than one pixel at a time, allowing the network to detect uncertain patterns in the image more efficiently. CNN is made up of three layers: an input layer, an output layer, and multiple hidden layers. CNN extracts features from an input image using a filter, which is an array of weights. Furthermore, as the number of channels increases, the height and width decrease. Finally, the generated column matrix is used to predict the output.

**Image Acquisition:**

Image acquisition is process of converting an analog image into digital form.

**Pre-processing:**

Pre-processing is a sequence of operations which improves quality of image and increase accuracy.

Noise Removal:- Remove noise from image, smoothening or reducing unwanted signals, minimizing or maximizing filter.

Binarization:- Converting gray scale or colored image to binary image i.e. 0’s and 1’s.

Morphological operation:- Increasing or decreasing image size because algorithm would expect the constant size.

**Segmentation:**

Extraction of individual character. From each row and column words are extracted using histogram.

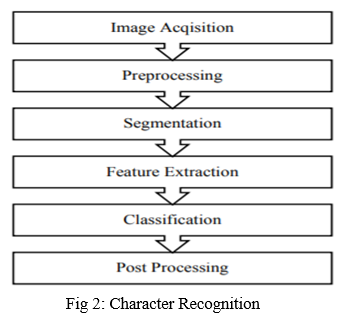
**Feature Extraction:**

Each and every character has features or group of rules or characteristics. The main aim of feature extraction phase is to extract the pattern which is most pertinent for classification. Histogram might be applied to extract the features of individual characters. These features are used to train the model.

**Classification:**

Here , the testing of input data starts. The testing data would pass all the above process and the varying probability are designed to the matching rules. The rule with highest probability is selected and label is recognuized.

**Phases:**

****

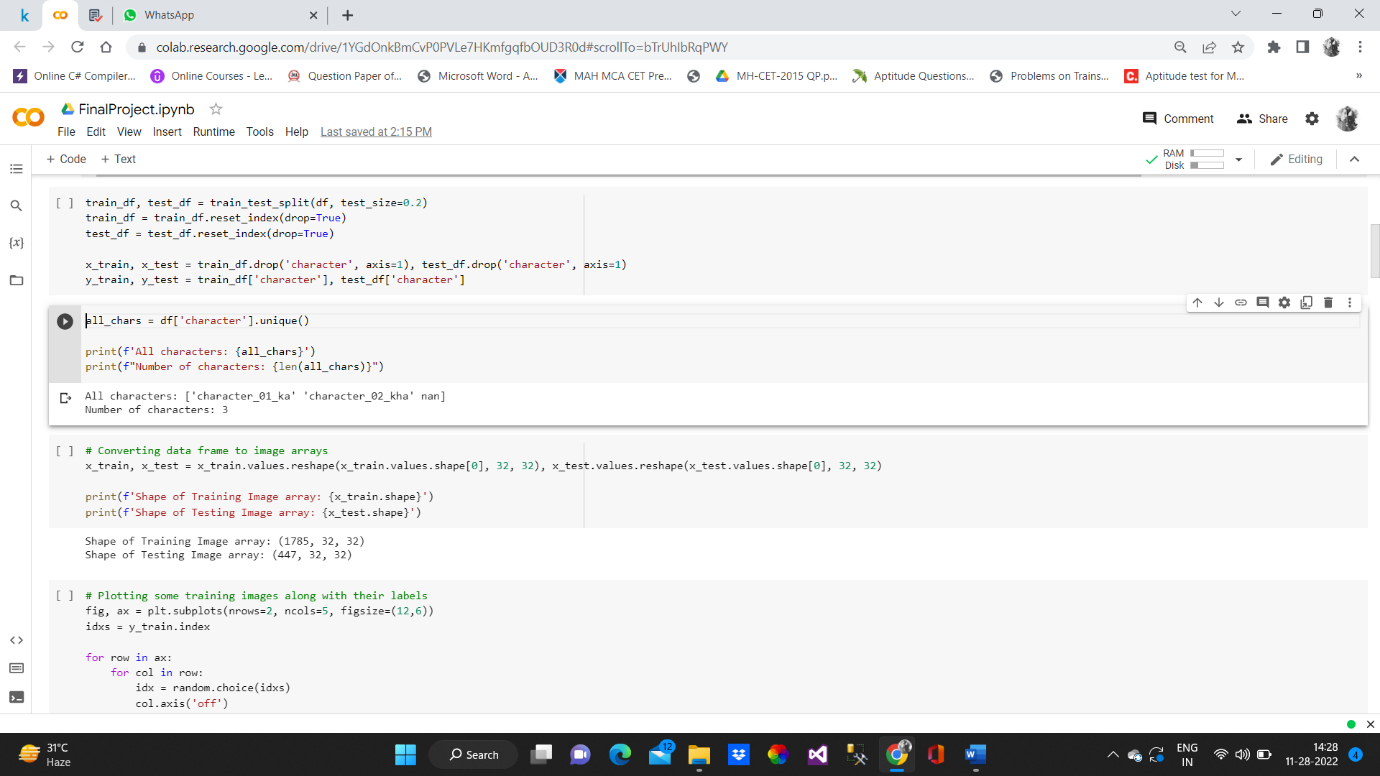
**About Dataset:**

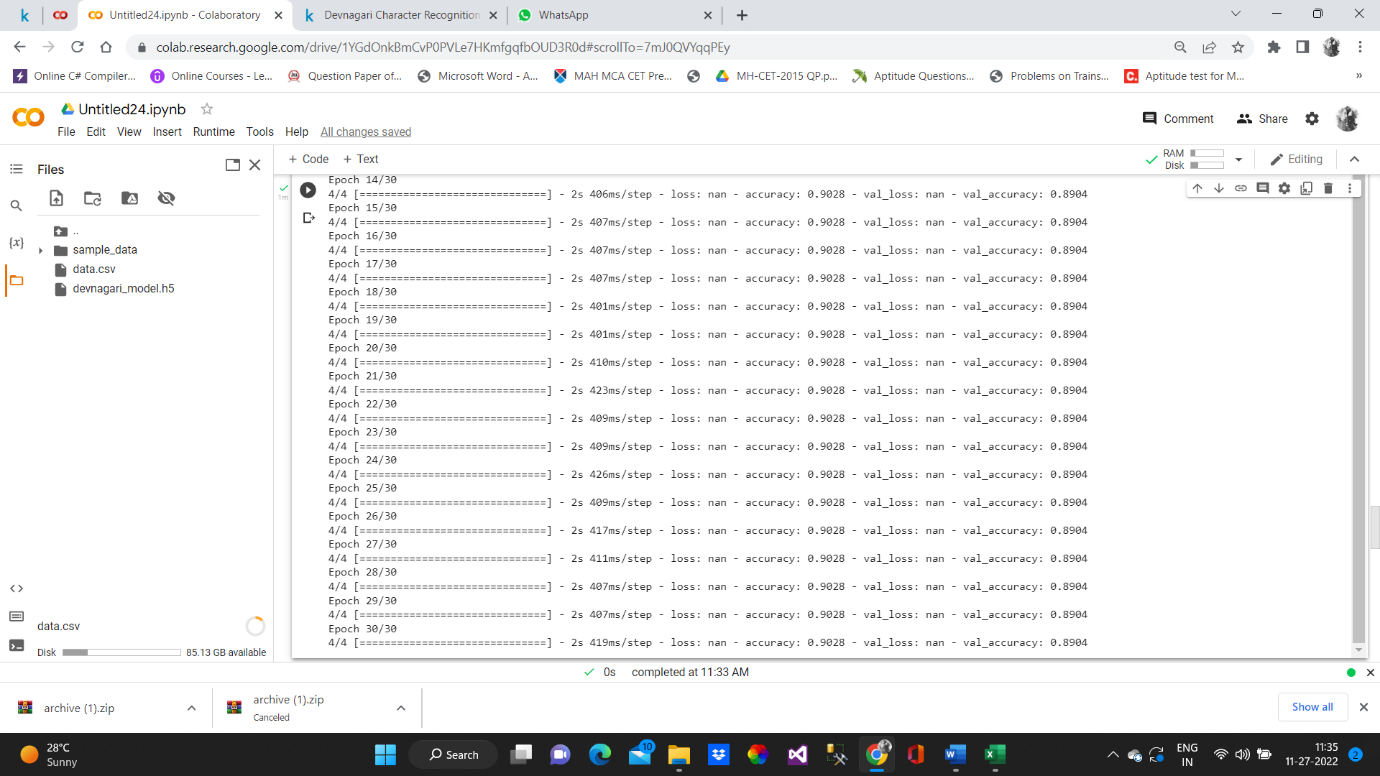
This is a dataset of Devanagari Script Characters. It consists of 92001 images [32x32 px] having 46 characters in it and consonants "ka" to "gya".

The CSV file is of the dimension 92001 \* 1025. There are 1024 input features of pixel values in grayscale (0 to 255). The column "character" represents the Devanagari Character Name corresponding to each image.

**Results:**

The results of our research show that deep learning technology achieved a maximum accuracy of 90.28% . The results of the experiment show that the CNN classifier is effective at recognizing handwritten digits.



****

**Prediction:  
**

**Conclusion:**

The results of our research show that by using CNN, we achieved a maximum accuracy of 90.28%. The most important problem in any recognition process is addressing feature extraction and correct classification approaches, which we attempted to do in our project. The algorithm that we used tries to address both factors and does so well in terms of accuracy and time complexity.

**Literature Review:** India is a multilingual country of around 135 crores (1.35 billion) population. Marathi, Hindi are the most popular languages written in devanagari script. Hindi is the national language of India and the third most spoken language of the world after Chinese and English. Hindi is used for documentation especially in indian states of New Delhi, Rajasthan, Uttar Pradesh, Madhya Pradesh, Himachal Pradesh, Uttarakand, Bihar, Cha ttisgarh and Haryana. So devanagari script is used in order to fill up various paper documents like bank cheques, envelops, application forms, railway reservation forms, answer sheets etc and also with the increase in popularity of internet, the number of websites hosted in devanagari has been increasing. So, there is a need for development of search engines which can search for sites/keywords provided in devanagari script. Handwritten character recognition of Indian script is a challenging task due to several reasons like huge number of characters, complex shape of the character and presence of modifiers.

Dr. Neeraj Pratap and Dr. Shwetank Arya of Gurukul Kangri Vishwavidyalaya said that , In the last two decades the different methods have been proposed by the Researchers for treating the problem of Devnagari character. But a lot of research is also required to handle the Challenges in Devnagari Character Recognition. It is hoped that this detailed discussion will be beneficial insight into various concepts involved, and boost further advances in the area. The accurate recognition is directly depending on the nature of the material to be read and by its quality.

**References:**

1. Review of Devanagari character recognition from past and future by Dr. Neeraj Pratap and Dr. Shwetank Arya of Gurukul Kangri Vishwavidyalaya.

<https://www.researchgate.net/publication/283510125_A_Review_of_Devnagari_Character_Recognition_from_Past_to_Future>

1. A review on research of Devanagari character recognition , authors: V.J. Dongre, Government polytechnic, Washim, Maharashtra and V.H. Mankar , Government Polytechnic, Nagpur, Maharashtra.

<https://arxiv.org/abs/1101.2491>

#### Handwritten Devanagari Character Recognition. Akhil Deshmukh, Rahul Meshram, Sachin Kendre, Kunal Shah, Department of Computer Engineering,Sinhgad Institute of Technology (SIT) Lonavala,University of Pune, India.https://www.ijert.org/handwritten-devanagari-character-recognition

1. Sonal Khare, Jaiveer Singh, Department of Com. Sci., Krishna institute of technology, Gaziabad.

<https://research.ijcaonline.org/volume121/number9/pxc3904600.pdf>

Code References:-

1. Akash, Student at self-Education,Delhi

<https://www.kaggle.com/code/aakash4235/devanagari-character-recognition>

1. Sahil Mani, student of Army Institute of technology, pune. <https://www.kaggle.com/code/devilmi/devanagari-character-recognition-1>
2. Suyog Dahal, Kathmandu, central development region, Nepal.

<https://www.kaggle.com/code/suyogdahal/devnagari-character-recognition-with-convolutions>

1. Peeyush kant Misra, developer at Cognizant technology solutions, Lakhimpur, Uttar Pradesh.

<https://www.kaggle.com/code/pkmisra/handwritten-letter-recognition>