MODI Charater Recognition using Convolutional Neural Networks

Aparna Shirkande [1], Alok Agarwal [2]

[1] Dept. of Electronics and communication, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India
kaleaparna5@gmail.com

[2] Dept. of Electronics and communication, Shri Jagdish Prasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India
alokagarwal26aaa@gmail.com

Abstract

Convolutional Neural Network also known as ConVNet or CNN algorithm is a Deep learning algorithm, which automatically detects the important features in an image without human supervision. The preprocessing required for ConVnet is much lower comparatively to other classification algorithm. The CNN successfully determines the spatial and temporal dependencies in an image through several filter. The CNN model has different layers that makes classification and feature extraction process easier. Thus CNN can be used for character recognition process. With help of Character recognition using CNN we can convert the Modi handwritten script into digitized English text and corresponding Devanagari character script. Modi script is used to write the Marathi language. Most of the ancient documents in Maharastra were written in Modi language. Hence the translation of the Modi script is necessary.In this paper the character recognition of the handwritten Modi script is performed using CNN algorithm. The pre-processing, training of CNN model are all written in python programming language in JupyterLab. The system developed not only translates the image character to the English but also display Devanagari corresponds of the Modi script predicted.The overall system accuracy is 96% for 46 character of the Modi script used.

Keyword: Preprocessing, CNN, Feature Extraction, Classification, MODI script.

I. Introduction:

The Modi script comes from the Nāgari script family and is a modification of the Balbodh style of the Devanagari script designed for continuous writing. Although Modi is based on Devanagari, it differs greatly from it in terms of letter forms, rendering behavior and orthography.The Modi script has 46 distinctive letters, of which 36 are consonants and 10 are vowels. Most of the documents in Modi are written by hand. The oldest document in the Modi script is from 1389 and is preserved at the BhāratItihāsSanshodhan Mandal (BISM) in Pune[5]. Most of the pre-Shivaji RajeBhonsle documents and correspondence are written in the Modi script.The Devanagari script is an important and widely used script in India. It is mainly used to write Hindi, Marathi, Nepali and Sanskrit. It also serves as an auxiliary script for other languages such as Punjabi, Sindhi and Kashmiri.

Character recognition is the process that allows computers to recognize written or printed characters, such as numbers or letters, and convert them into a form that the computer can use[2].

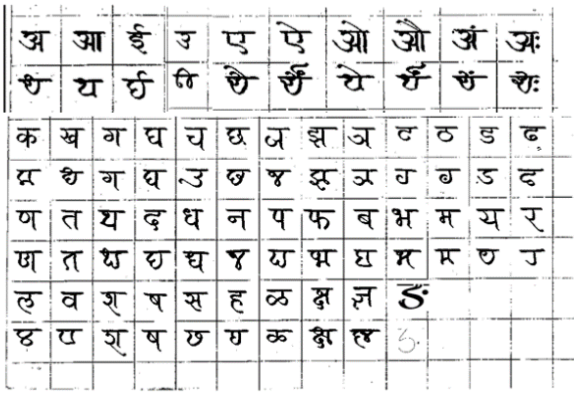


Fig.1: Dataset

Although OCR has been considered a solved problem, there is one key component to it, handwriting recognition. (Handwritten OCR) or Handwritten Text Recognition (HTR), which is still considered a challenging representation of the problem. The wide variation in handwriting styles between people and the poor quality of handwritten text compared to printed text present significant obstacles to conversion to machine-readable text[14].

Character recognition steps:

1.Image acquisition: The input Modi character image is given as the input to the system .

2.Preprocessing:The goal of pre-processing is to improve image data that suppresses unwanted distortion or improves some image properties important for further processing

3. Feature Extraction:The feature extraction phase is used to extract the most important information from the text image, which helps us recognize the characters in the text.In the feature extraction phase, each character is represented as a feature vector that becomes its identity[11].

4.Classification: Image classification is the process of categorizing and labeling groups of pixels or vectors within an image based on specific rules. A categorization law can be designed using one or more spectral or textural characteristics.

5.Prediction: The model trained is used on the input image to predict the result or the prediction of the character with its' label.

II. Literature Review

Several previous worked has been done on character recognition of Modi script using different classification algorithm and achieved different accuracy level. In the paper "Handwritten Character Recognition of MODI Script using Convolutional Neural Network Based Feature Extraction Method and Support Vector Machine Classifier" proposed by Solley Joseph,Solley Josephuses CNN autoencoder as a feature extractor and an SVM based classifier gives out the accuracy of 99.3 %[1]. In 'Recognition of Characters in Indian Modi Script' paper proposed by S.S. Gharde, R.J. Ramteke,Moment Invariant and Affine Moment Invariant are the two techniques which are used for extracting the features from the handwritten segregated samples. Identification and recognition is done using machine learning technique. Among the variants of machine learning techniques, support vector machine is used as a classifier. Linear kernel function is applied in this support vector machine while performing classification. The paper result to give recognition rate of 89.72%[4].

'Recognition of handwritten modi digits and characters by using deep learning algorithm' by JidnyasaKondhare, Prof.V.A.Yaduvanshi, Ruchira Patil, Radhika Kaldatedeveloped MODI Script Character Recognizer System (MSCR) using Convolutional Neural Network (CNN) algorithm and vgg16 .The vgg16 algorithm shows the training accuracy of 99.73%,CNN algorithm achieved the training algorithm accuracy of 73.93%[6].Vishal Pawar, Deepak Wadkar, ShivkumarKashid, PremrajPrakare, Vinayak More, Prof. S.A. Babar proposed 'MODI Lipi Handwritten character Recognition using CNN and Data Augmentation Techniques' paper in which CNN model was implemented and it gave the accuracy of 91.62% [7].

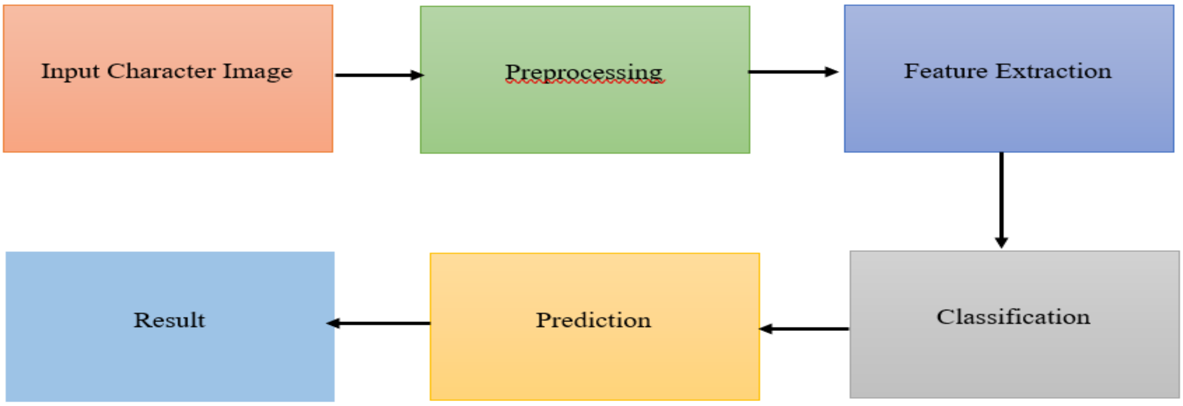


Fig.2 : Character Recognition Steps

Recognition of Off-line Modi Script : A Structure Similarity Approach, Proceedings by A.S. Ramteke, G.S. Katkar, used different methods and patterns of classification, pre-processing and segmentationrecognition techniques such as Kohenen Neural Network and back-propagation neural network. An attempt is made to apply the measured structural similarity approach off-linerecognition of handwritten Modi characters. SSIM performance was found to be 91to 97 percent[3].Feature extraction method using BackpropagationNeural network and structure similarity has been implementedfor MODI script recognition and accuracy93.5% was achieved [8].An implementation of seven Hu feature extraction with a Euclidean Distance classifier for MODI feature recognition resulted in an accuracy of 82.6%[9]. For MODI script character recognition, a feature extraction method using empirically determined heuristics was implemented and achieved an accuracy of 91.2%[10].

III. Methodology

1.Preprocessing

Preprocessing Isan critical step
to easy image records before it is prepared for
use in a pc vision model. There are both technical
and overall performance reasons why
preprocessing is important. The steps performed
are as follows:

1.Binarization:Before binarization colour image is converted to greyscale. Binarization is converting an image from a grayscale image to a black and white image (ie 0 and 255 pixels). This can be achieved through a process called thresholding. So using thresholding we can get binary images. Now, thresholding can be defined as a process in which each pixel is converted to 0 or 255 depending on whether its value is greater or less than the threshold value. If the pixel value is greater than the threshold, it is converted to 255, otherwise it is converted to 1.

2. Inversion:When you invert a black-and-white photo, the white pixels become black pixels and vice versa.It is needed for dilation and erosion stage.

3.Dilation and erosion:Dilation and erosion are basic morphological processing operations that produce contrasting results when applied to grayscale or binary images. Dilation is the opposite process where regions grow out of their boundaries. Erosion involves removing pixels eating away at the edges of an area.

4. Resizing: Image resizing is needed as the input image dimension can be of any dimensions . It is needed for the proper prediction of the image that the recognition of the image size is as mentioned in the trained model.

5.Noise removal: Noise in an image is the unwanted distortion that degrades the quality of the image. It is hence essential to remove the noise and enhance the feature of the character in an image.

2.CNN

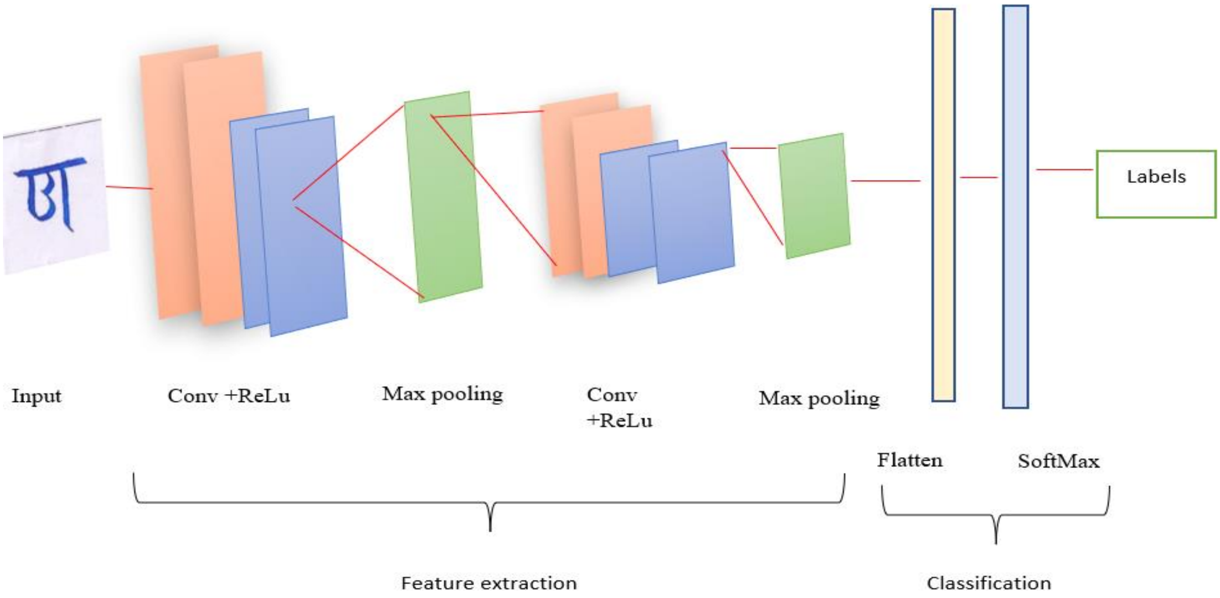


Fig.3 : Architecture of CNN

In deep learning, a convolutional neural network (CNN) is a class of artificial neural network most.A convolutional neural network is a feedforward neural network that is generally used to analyze visual images by processing data with a grid-like topology. It is also known as ConvNet. A convolutional neural network is used to detect and classify objects in an image. CNNs use a mathematical operation called convolution instead of general matrix multiplication in at least one of their layers. They are specially designed to process pixel data and are used in image recognition and processing. They have applications in image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series.

There are multiple hidden layers like the convolution layer, the ReLU layer, and pooling layer, that perform feature extraction from the image. Finally, there's a fully connected layer that identifies the object in the image.

3.CNN layers

3.1. Convolutional layer

Convolutional layers convolute the input and pass its result to the next layer. It is similar to the response of a neuron in the visual cortex to a particular stimulus. Each convolutional neuron processes data only for its receptive field. Convolutional layer is the basic building block of CNN[12]. It carries the main part of the computing load of the network.This layer performs a dot product between two matrices, where one matrix is a set of learnable parameters otherwise known as a kernel and the other matrix is a bounded portion of the receptive field. The core is smaller in space than the image, but more in depth. This means that if the image is composed of three (RGB) channels, the height and width of the kernel will be spatially small, but the depth extends to all three channels. During the forward pass, the kernel moves along the height and width of the image to create an image representation of this receptive field. This creates a two-dimensional representation of the image known as an activation map that provides the response of the kernel at each spatial position of the image. The sliding size of the kernel is called the step.

Mathematically:

If size of image: MxM

If size of filer: KxK

Then (MxM)*(KxK)= (M-K+1)x(M-K+1)

3.2: ReLu

In this layer, we remove every negative value from the filtered image and replace it with zero. This function is only activated when the input of the node is above a certain quantity[13]. So when the input is below zero, the output is zero. However, when the input exceeds a certain threshold, it has a linear relationship with the dependent variable. This means that it is able to speed up the training data set in a deep neural network faster than other activation functions - this is done to avoid zero summing.

3.3. Pooling Layer

A pooling layer is added after the convolutional layer(s) as seen in the CNN structure above. It resamples the output of convolutional layers by shifting a filter of a certain size with a certain step size and computing the maximum or average of the input.

1. Max Pooling: In max pooling, the maximum value is taken from the patch function map value group.

2. Minimum pooling: In this type of pooling, the minimum value is taken from a patch in the feature map.

3. Average Pooling: Here the average of the values is taken.

Consider object map dimensions as HxWxD.

Let H- height of object map

W- width of object map

C- the number of channels on the function map

F-filter and S- step length

Then the output image will be:

[(H-F)/S] +1 x [(W-F)/S] + 1 x C

3.4. Flattening and fully connected layer

Flattening involves taking the pooled feature map that is generated in the pooling step and transforming it into a one-dimensional vector. The reason we transform the concatenated feature map into a one-dimensional vector is that this vector will now be fed into the artificial neural network. The purpose of a fully connected layer in a convolutional neural network is to detect certain features in an image. More precisely, each neuron in a fully connected layer corresponds to a specific feature that may be present in the image. The value that the neuron passes to the next layer represents the probability that the element is contained in the image.

3.5. SoftMax

Softmax activation functionwill generate the probabilities i.e. if there are n classes, then for a given input image, it will generate probabilities for each of the n classes and the one with the highest probability is the output.

$$\sigma(\vec{z}_i) = \frac{e^{-z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

Where

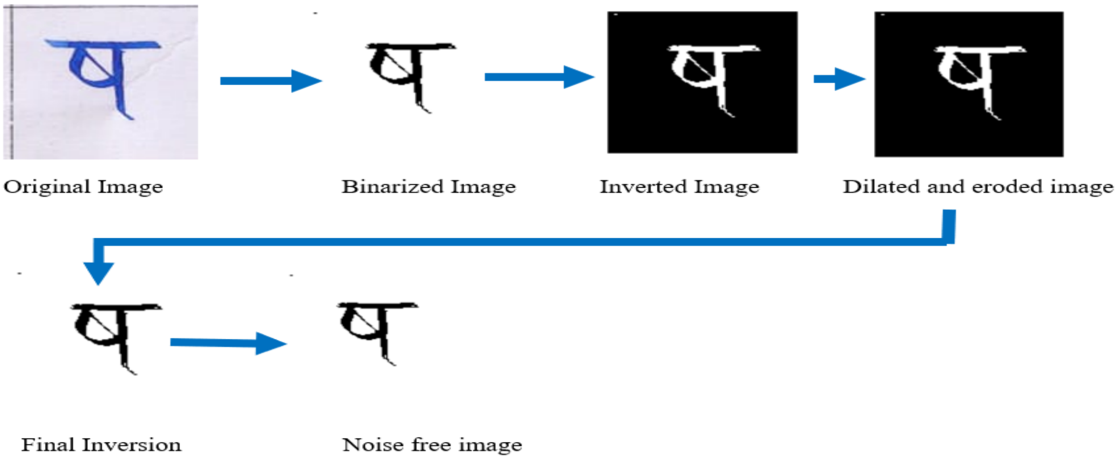
σ = softmax

z_{i}= input vector

K= number of class in multiclass classifier.

IV. OUTPUT

Fig.4: Output of Pre-processing Stage



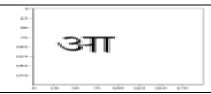
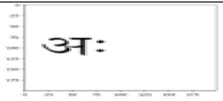
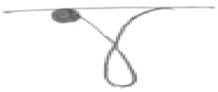
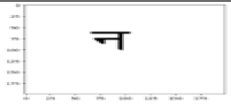
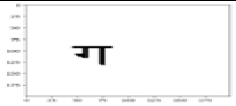
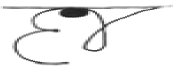
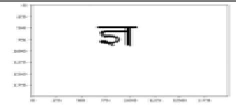
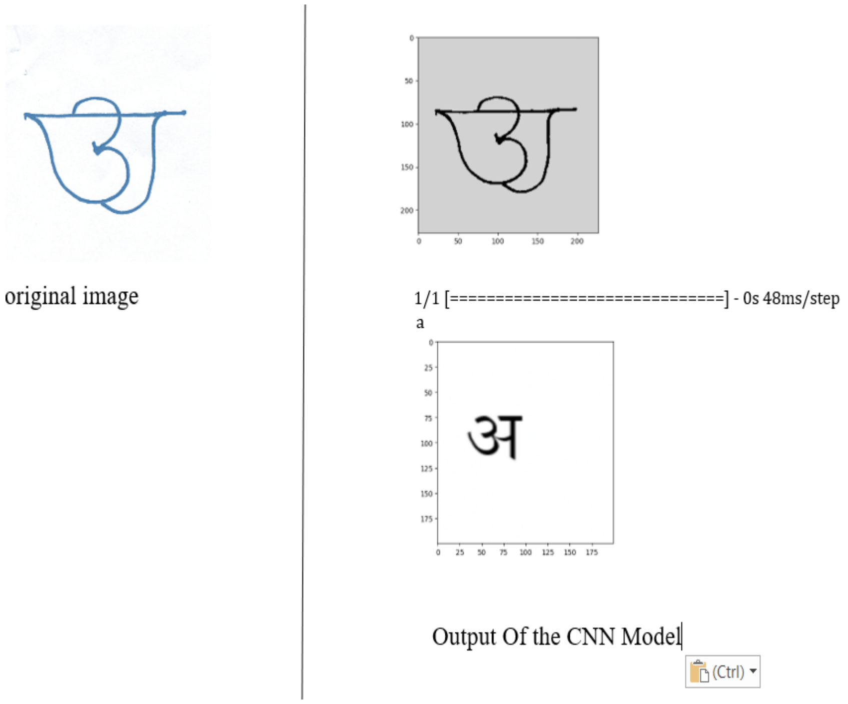
Input Image	Expected	Predicted
	आ	
	अः	
	न	
	ळ	
	ज्ञ	

Fig.5: Output of the training model



original image

```
1/1 [==============================] - 0s 48ms/step
a
```

Output Of the CNN Model

Fig.6: Chart of Expected vs Predicted output

V. Experimental setup:

The experiment was performed on the JupyterLab workspace (anaconda navigator) which is the open source platform.The total dataset consists of 4151 images for 46 characters of the MODI script. The training model has been trained on the 3323 images and testing dataset comprises of 828 images.Afterpreprocessing stage, the character image uses the trained model to perform prediction. Data Augmentation is performed on training dataset using Keras inbuilt function imageDataGenerator. The CNN model is designed and compiled and trained on the training dataset using TensorFlow library functions. The accuracy model and the loss model graph is achieved.

Model accuracy is defined as the number of classifications the model correctly predicts divided by the total number of predictions made. It is a way of evaluating the model's performance, but certainly not the only one.Loss is a value that represents the sum of errors in our model. It measures how well (or poorly) our model is doing. If the errors are high, the loss will be high, which means the model is not doing a good job. The system model gives the accuracy of the 96% for 7.05% loss .

Accuracy as well as loss is shown against the epoch. An epoch is when all the training data is used at once and is defined as the total number of iterations of all the training data in one cycle for training the machine learning model. Another way to define an epoch is the number of passes a training dataset takes around an algorithm. One pass is counted when the data set has done both forward and backward passes. The total number of epoch used were 200. Hence overall accuracy is 96%.

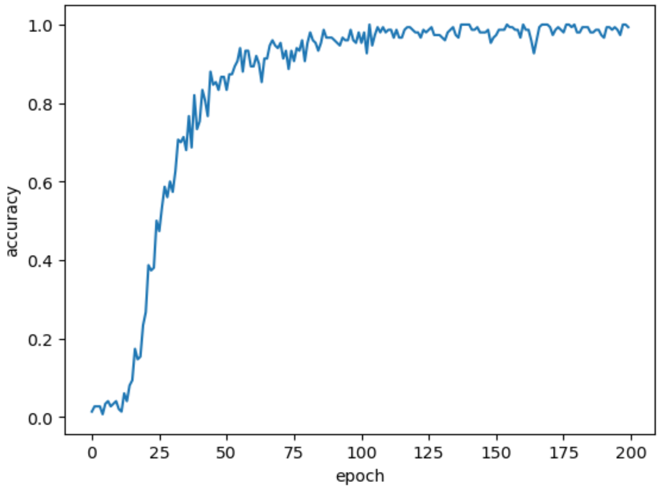


Fig.7 : Accuracy model

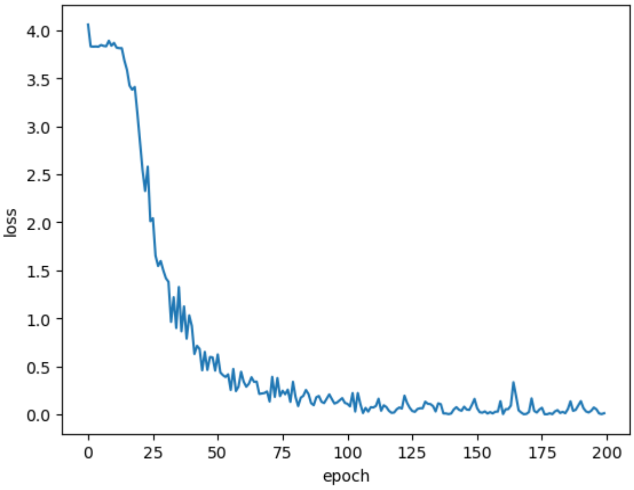


Fig.8 :loss model

VI. Conclusion

In the proposed system, we have performed handwritten character recognition of the MODI script using CNN.the experiment was performed on the JupyterLab workspace and the code is written in the python Programming language . The preprocessing stage is used to clear the character image of the handwritten MODI script that is to be predicted, which will be helpful to ease out the process of the prediction.MODI handwritten character dataset of total 4151 images are used for total of 46 characters out of which 3323 images are used for training and 828 images are used for testing. The simple libraries in Python are used such as Keras and Tensorflow .The CNN model is designed using Keras library in the python. The model is designed, compiled and fitted for training the Convolutional Neural Network algorithm.The loss is categorical crossentropy and the optimizer used is adam. The total number of the epoch used over the CNN model is 200 and the step per epoch is 50. The graph of the accuracy and loss against the epoch are successfully acquired.The system is successful to achieve the overall accuracy of 96% and can be used for predicting the MODI character and display its digital English text and corresponding Devanagari character image.

References

[1] Solley Joseph,Jossy George, "Handwritten Character Recognition of MODI Script using Convolutional Neural Network Based Feature Extraction Method and Support Vector Machine Classifier",2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP) | 2020 IEEE | DOI: 10.1109/ICSIP49896.2020.9339435

[2] K. Sadanand, P.L. Borde, M. Ramesh, Y. Pravin, Offline MODI Character Recognition Using Complex Moments, Procedia Comput. Sci. 58 (2015) 516–523. https://doi.org/10.1016/j.procs.2015.08.067.

[3]A.S. Ramteke, G.S. Katkar, Recognition of Off-line Modi Script : A Structure Similarity Approach, Proceedings- Int. J. ICT Manag. 1 (2013) 12–15.

[4] S.S. Gharde, R.J. Ramteke, Recognition of characters in Indian MODI script, Proc. - Int. Conf. Glob. Trends Signal Process. Inf. Comput. Commun. ICGTSPICC 2016. (2016) 236–240. https://doi.org/10.1109/ICGTSPICC.2016.7955304

[5] S.L. Chandure, V. Inamdar, Performance analysis of handwritten Devnagari and MODI Character Recognition system, Int. Conf. Comput. Anal. Secur. Trends, CAST 2016. (2017) 513–516. https://doi.org/10.1109/CAST.2016.7915022.

[6] JidnyasaKondhare, Prof.V.A.Yaduvanshi, Ruchira Patil, Radhika Kaldate,RECOGNITION OF HANDWRITTEN MODI DIGITS AND CHARACTERS BY USING DEEP LEARNING ALGORITHM,2022 JETIR August 2022, Volume 9, Issue 8,ISSN-2349-5162

[7]Vishal Pawar, Deepak Wadkar, ShivkumarKashid, PremrajPrakare, Vinayak More, Prof. S.A. Babar ,MODILipi Handwritten character Recognition using CNN and Data Augmentation Techniques ,International Research Journal of Engineering and Technology

(IRJET),Jun 2022,Volume: 09 Issue: 06,p-ISSN: 2395-0072.

[8] A.S. Ramteke, G.S. Katkar, Recognition of Off-line Modi Script : A Structure Similarity Approach, Proceedings- Int. J. ICT Manag. 1 (2013) 12–15

[9] P.L. Borde, R.R. Manza, S. Kulkarni, P.L. Borde, R.R. Manza, P. Yannawar, Analysis Of Orthogonal Moments For Recognition Of Handwritten MODI Numerals, J. Sci. Technol. 4 (2015) 36–43.

[10] R.K. Maurya, S.R. Maurya, Recognition of a Medieval Indic- ‘ Modi ’ Script using Empirically Determined Heuristics in Hybrid Feature Space, (2018).

[11] U. Pal, B.B. Chaudhuri, Indian script character recognition : a survey, Pattern Recognit. 37 (2004) 1887–1899. https://doi.org/10.1016/j.patcog.2004.02.003.

[12]WeiXiong , Bo Du, Lefei Zhang, Ruimin Hu, Dacheng Tao "Regularizing Deep Convolutional Neural Networks with a Structured Decorrelation Constraint ” IEEE 16th International Conference on Data Mining (ICDM) , pp. 3366–3370, 2016.

[13]]“Study on the OCR of the Devanagari script using CNN”Prof. Aparna. S. Shirkande , Sakshi. S. Sawant , Neha. V. Shinde ,Sharanya. S. Rao, International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue IX Sep 2022

[14]Prasad Chavan, SuyogSankpal, AkshaySonawane, Shahid Shaikh, Prof. Anup Raut, "Handwritten Devanagari Optical Character Recognition", International Journal of Innovative Research in Computer Science and Technology, ISSL-2347-5552, Volume 2, Issue 2, March 2022