***Handwritten Devanagari Character Recognition***

# Sanjay Jaras DSC680 Applied Data Science GitHub Portfolio URL: <https://sanjayjaras.github.io/>

# Domain and Background

In the past few years, deep neural networks are getting highly used in computer vision and machine learning tasks such as regression, segmentation, classification, detection, pattern recognition, etc. because of their outstanding performance. Devanagari script also called Nagari script is mostly used in the Indian subcontinent. This script is based on the ancient Brahmi script. Devanagari script is the most used Indian script. The Devanagari script is composed of 47 primary characters including 14 vowels and 33 consonants It is the fourth most widely adopted writing system in the world being used for over 120 languages. Some of the languages using Devanagari script as a writing system are Sanskrit, Hindi, Marathi, Pali, Kashmiri, and Nepali. This script is used by more than 500 million people worldwide.

**References:**

[1] Devanagari Handwritten Character Dataset Data Set - Center for Machine Learning and Intelligent Systems <https://archive.ics.uci.edu/ml/datasets/Devanagari+Handwritten+Character+Dataset>

[2] Classifier comparison on Devanagari recognition - Prashanth Aditya Susarla - <https://www.kaggle.com/spadix/classifier-comparison-on-devanagari-recognition>

[3] Handwritten Devanagari character identification with ResNet - Akshay Chougule - <https://medium.com/@akshaychougule/handwritten-devanagari-character-identification-using-resnet-b90894b42c4d>

[4] Devanagari - <https://en.wikipedia.org/wiki/Devanagari>

[5] Deep Learning based Character Classification using Synthetic Dataset - Krutika Bapat - <https://learnopencv.com/deep-learning-character-classification-using-synthetic-dataset/>

[6] Devanagari Character Set - Rishi Anand - <https://www.kaggle.com/rishianand/devanagari-character-set>

[7] Intro to Computer Vision-Indian National Language - Shruti Iyyer - <https://www.kaggle.com/shrutimechlearn/intro-to-computer-vision-indian-national-language>

[8] Devanagari Character Dataset - Ashok Pant - <https://www.kaggle.com/ashokpant/devanagari-character-dataset>

[9] Image classification - <https://www.tensorflow.org/tutorials/images/classification>

[10] Recognizing Handwritten Digits with Scikit-learn -

[Manthan Bhikadiya](https://manthan-bhikadiya.medium.com/?source=post_page-----90ca6e2471ed--------------------------------) - <https://medium.com/codex/recognizing-handwritten-digits-with-scikit-learn-90ca6e2471ed>

# Data Understanding

This dataset is obtained from Machine Learning Repository([Center for Machine Learning and Intelligent Systems](http://cml.ics.uci.edu/))[1]. This is an image database of Handwritten Devanagari characters. There are 46 classes of characters with 2000 examples each. The dataset is split into a training set(85%) and a testing set(15%). So in total, there are 92000 images. The training dataset has 78200 images and the Test dataset has 13800 images. Each image is of 32x32 pixels with one channel. The actual character is centered within 28 by 28 pixels, and padding of 2 pixels is added on all four sides of the actual character. The 46 classes contain 10 classes for digits and 36 for other characters.

# Research Questions:

Every character is written with several variations while writing the Devanagari as a writing system. Devanagari and in turn Hindi is the national language of India and Nepali is the national language of Nepal and adopted by more than 500 million people worldwide. Contrary to wide adaptation, not much of the work happened with the Devanagari script as compared to other scripts like English. The handwritten character recognition will help in preserving the ancient document written in the Devanagari script. There are other applications of offline handwriting recognition like reading postal addresses, bank check amounts, and forms. Creating digital libraries, allowing the entry of image textual information into computers by digitization, image restoration, and recognition methods.

# Method:

Character recognition is classified into two categories namely online and offline character recognition. I am planning to do offline handwritten character recognition. I will be using Convolution Neural Network with dense layers to create a machine learning model that will recognize the handwritten characters. I am using CNN as it is very good while working with images. If time permits, I will also try other models by using algorithms.

# Potential Issues:

Handwriting recognition in Devanagari script is a challenging problem due to the inherent subtleties in the scripts, cursive nature of the handwriting, and similar shape of the characters.

# Concluding Remarks:

With a well-trained model for recognizing handwritten Devanagari characters, it can be further extended to train the model to read the whole Devanagari script. This will help in different applications like preserving the ancient documents, reading addresses, bank checks, government applications filled/submitted in Devanagari script.