

Deep Learning Model to Revive Indian Manuscripts

Puran Bhat¹, Kannagi Rajkhowa²

¹Chandigarh University, University Institute of Engineering,
NH-05, Ludhiana- Chandigarh State Highway, SAS Nagar, Punjab, 140413, India
21BCS10847[at]cuchd.in

²Chandigarh University, University Institute of Engineering,
NH-05, Ludhiana- Chandigarh State Highway, SAS Nagar, Punjab, 140413, India
21BCS10790[at]cuchd.in

Abstract: Sharda is an ancient classic Kal script of Kashmir. It is said to have evolved from the 'Brahmi' also some 1200 years ago. It was much used in undivided Kashmir, North, and Western India (Gilgit), etc, Punjab, Himachal Pradesh, and even up to Central Asia. Devanagari a combination of two words Dev and Nagri (Script of God) is the mother of more than 100 languages spoken in undivided India (Hind, Kabul, Sind, Pakistan, Barhma, Lanka, and Bhutan, etc), comprises 13 vowels, 36 consonants, 10 digits. When a vowel is added to a constant the alphabet here gets modified. The Sharda script also supports various compound characters. Due to this recognition is a challenging task. This paper deals with the deep learning module to revive this Indian Manuscript Sharda. Devanagari is an ancient Indian alphabet. This script can be used in Sanskrit, Hindi, Marathi, Bhojpuri, Nepali, Konkani, Sindhi, Marwari, Maithiri language, Pali etc. It is used to write more than 100 language spoken in India and Nepal, including The word "Devanagari" is a combination of two words meaning God and "Nagari" meaning house. Therefore, the word "Sanskrit" means "God's writing". It contains 49 basic letters, 13 sounds and 36 numbers and 10 digits. Also, letters are changed when vowels are added to consonants. The script also supports multiple character combinations. Recognition of the Devanagari script is therefore a challenge and has been an area of research to this day. Researchers have used different types of analysis such as neural networks, support vector machines, particle swarm optimization techniques, and genetic algorithms. This article discusses different classification studies for Devanagari character knowledge.

Keywords: Sharda, Devanagari, Deep Learning

1. Introduction

One of the most prominent and famous classic temple universities in the Indian Subcontinent was "Sharda Peeth" between the 6th -12th centuries C.E. Sharda Peeth means the seat of Goddess Sharda, the Goddess in charge of learning, wisdom, and power. Sharda is the Kashmiri name of the Hindu goddess Saraswati. LalitadityaMuktapedia between 724 C.E to 760 C.E is said to have commissioned the Peetha. The Sharda Peeth was famous for its library. Distant and local scholars visited this temple university to access its texts which were mostly in the Sharda Script. The grammar for many languages was made and studied here. It was the best site for pilgrimage also attracting devotees from as far as present-day Bengal. History reveals that the first convocation address of the Sharda Peeth University was delivered by Kataliya also known as Chanakya. Now it is a Kashmiri ruined temple at 34 48 latitudes and 74 14 longitudes.

Sharda Peeth was associated with Buddhist scholars, Kamraja, Sambota, Richen Zangpa, Kalhan Pandita, and Adi Shankara. It is located at village Shardi (Sharda), a small Tehsil in Neelam Valley (Krishen Ganga) District in Azad Kashmir Pakistan; the "Mount Harmukh" at an altitude of 6499 feet on the confluence of Krishna Ganga and Madhumati. A parallel Sharda Temple has been constructed in Teetwal District Kupwara just opposite it near LOC.

Sharda Peeth played an important role in the development and popularisation of the Sharda Script in Northern India causing the script to be named after it and Kashmir to acquire the moniker, Sharda Peeth meaning Country Of Sharda.



Figure 1: Sharda script

Sharda script in India and its adjoining central Asian countries have developed from Brahmi and early LBrahmic inscriptions. It dates back to the "Morean Period). Sharda Script was much in use in Kashmir where Sanskrit, Urdu, Hindi, Kashmiri, and Sharda itself were written in this script but in North Western India (Gilgit, etc), Punjab, Himachal Pradesh, and even in Central Asia.

Volume 12 Issue 4, April 2023

www.ijsr.net

Licensed Under Creative Commons Attribution CC BY

Sharda Script enjoyed a foremost position among all Indian Scripts. It is and was an excellent, ancient alphabet of Kashmir. Almost all the ancient Sanskrit texts, Literature, Hymns, Sehisternamas,(1000 names), Alminaics, Palmistry, Tantra, Yoga, Ayurveda, and Athervveda in Kashmir were written in the script.

The earliest Sharda inscription on stone slabs (774) A.D. was discovered at village Hund at Attock in West Pakistan. The earliest phase is represented by inscriptions and coins of the 8th to 10th centuries. The second by those of the 11th - 14th centuries and the third and final by the epigraphics and literary records of the 14th and subsequent centuries.

Devnagri Script Origin

Devanagari is a combination of two Sanskrit words: deva meaning god, Brahman or celestial body, and Nagari meaning city. The name can mean "texts of the city", "heavenly/sacred texts of the city" or "[texts] of the city of the gods or priests. The Nagari or Devanagari alphabet is derived from the eastern Gupta alphabet called Nagari, which first appeared in the 8th century. The alphabet began to resemble the modern Devanagari alphabet in 10th AD and began to replace Siddham around 1200. It can be traced back to the Brahmi script. It has a Brahmi sound system. It was built between the 1st and 4th centuries AD. Brahmi script developed during the Vedic period. During the Buddhist and Jain periods, the Brahmi script spread to many parts of the subcontinent. It is known as the Bharati script and has become an essential part of the Bhagwat Gita. This Bharati script was changed to Gupta script, then Nagari script and finally Devanagari script during Gupta period.

The aim of the program is to help students and readers learn about scripture and Sharda script and to encourage them to explore and understand Indian wisdom in ancient times. Brahmi Script evolved during Vedic times. It was also named Bharti script and became a permanent feature in Bhagwat Gita. This Bharti to Gupta script during the Gupta period then to Nagri and ultimately to Devnagri Script as is shown below.

अ आ इ ई उ ऊ
 ऋ ॠ लृ लृ
 ए ऐ ओ औ
 क ख ग घ ङ
 च छ ज झ ञ
 ट ठ ड ढ ण
 त थ द ध न
 प फ ब भ म
 य र ल व
 श ष स ह

Figure 2: Devanagari script

2. Pattern Recognition Model

Pattern recognition works by performing various types of tasks. Out of these tasks, one of the prominent tasks is the Feature Extraction from the dataset or the data image. The

next important task is the stage in which patterns are recognized. This stage uses the data of the feature extraction stage to allocate a character that is unknown to one of its prominent feasible classes. We use any distant classifier/character to identify this feature of extraction. This recognition of the characters goes under various steps which are Acquisition, Preprocessing, Segmentation, Feature extraction, Classification, and Post-processing. The version is shown in the diagram below. In character recognition much significant research activities have emerged in past few decades. To make the character into a machine-understandable language from humans understandable we use handwritten Character recognition. Hand-written character recognition provides the solution for handling substantial volumes of information naturally. For it, various deep activities have been made on character recognition. It is also used in scientific and business applications. The other recognition model is Optical character recognition also known as OCR. It is also used in the management and recognition of characters that are handled optically instead of the already prepared ones. OCR works by recognizing the character optically and automatically by the Pc and is examined optimally and then examined by digitized. It is the most interesting as well as a challenging method but it also has different applications. It can be used widely in computer procedures and can enhance the interface between man and machine.

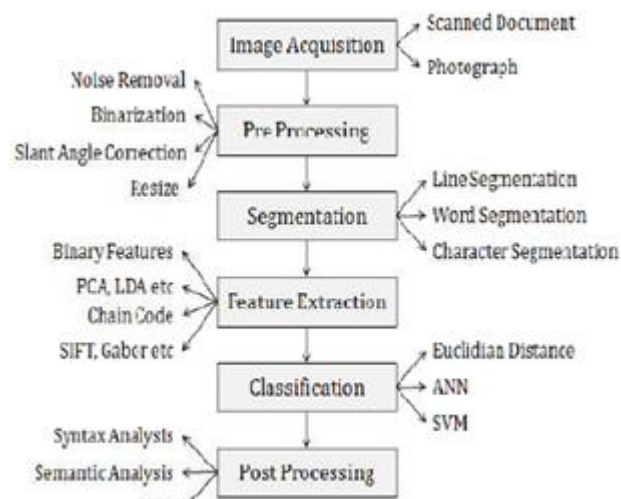


Figure 3: Pattern Recognition Model

3. Handwriting Character Recognition Methodology

This methodology includes feature extraction techniques and classification techniques.

3.1 Feature extraction

The characters and features which are essential are captured in this feature extraction. It plays an important role and is a very essential stage. The most important stage is selecting a feature from the extracted feature. It reduces the process of computation and thus the time taken in recognition will also reduce. When the time taken by the recognition reduces it enhances the performance of the system. Features are obtained from the arrangement of the points which constitute

the matrix (according to C.V Suen). As compared to the topological features, these features are easily dejected. It has high as well as low complexity. Various techniques are used in feature extraction which are moment, zoning, projection, histogram, n-topples, crossing, and distance.

3.1.1 Statistical features

The statistical distribution of the pixels and their characteristic measurement of pattern comprises the statistical feature. These are those measurements that represent various characters. Some of them are shape, size, internity, etc. The other various statistical features include major axis length, minor axis length, centroid, convex hull, convex area, convex image, bounding box, orientation, eccentricity, filled image, filled area, extent, extrema, etc.

3.1.2 Methods Used to extract statistical features

The various method used in the extraction of the statistical features is a moment, zoning, projection histogram, n-topple, crossing, and distance. These are the measures of some sort of inherited.

- 1) Moment: - In this case, character moment is utilized that is the moment of different points present in any character is utilized as a feature.
- 2) Zoning: - In the zoning the characters are divided into various zones. Based on these various zones its density is calculated. These zones are termed overlapping and non-overlapping zones. Density is calculated by using the formula number of object pixels in each zone divided by the total number of pixels.
- 3) Projection histogram: - projection histogram can be vertical, horizontal or diagonal (left or right). In the specified character it gives the number of black pixels in the vertical and horizontal directions.
- 4) N-topple: - This method helps in providing the number of important proportions of the pixels. It works on the position of black and white pixels. These positions are considered a pixel feature.
- 5) Crossing and distances: - In this type, features of the characters are counted and distance is measured at a certain angle in a certain direction by using a vector.

3.2 Structural features

The image description of its topological and geometric characteristics constitute the structural feature. Structural features are of various types. Some are listed below;

horizontal and vertical lines, endpoints, crossing points, curves at the top and bottom, curve opening direction, and counter numbers. These are very helpful in character recognition. various methods are used in the extraction of structural features. Some are listed below:

Geometrical and topological features: When the type of representation takes place it decodes some of the prominent knowledge about the object and sorts those components which make that object. It can further be distinguished into four subgroups.

- a) Extracting and counting topological structure: - Features which are distinguished in it are aslines, curves, spleens, extreme points, maxima, minima, etc.

- b) Measuring and approximating the geometrical Proprieties: By using geometric quantities characters are recognized such as the ratio between height and width of the bounding box of character, the relative distance between the last point and last y-min, the distance between the two points, upper and lower masses of words and word length curvature or change in curvature.
- c) Coding: Coding is used by mapping strokes of character in two dimensions. An example of which is the Freeman chain code.
- d) Graphs and trees: Topological primitives such as strokes, holes, cross points, etc are first used. These are represented by relational graphs.

3.3 Neural Network Approach

Compared to the structural approach and statistical approach, it is the neural network approach is used for better recognition of handwritten character recognition. It has the distinctive inherent property of reconciling learning. Neural networks can automatically deal with the peculiarities of different handwriting styles. Some of the neural network pattern recognition are Feedforward, networks, back propagation networks, hope field models, and self-organizing networks.

3.4 Classification

The decision-making stage of handwritten character recognition with homogeneous characteristics is classification. Based on stored features of the features space, i.e. structural and global features, it can be carried out.

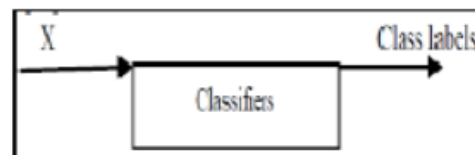


Figure 4: Classifier Technique, where X is the eigenvector.

Template matching, Statistical techniques, Structural techniques, and artificial are some classification methods.

- Template matching: This is the simplest form of pattern recognition. The given pattern that is to be recognized is compared with the stored pattern. Style and size are ignored while matching the pattern.
- Statistical method: In this task, the category of the given pattern belongs. By making observations and measurement process, a set of numbers is prepared, which is to prepare a measurement vector.
- Structural method: It is good for determining handwritten text. In this type, the pattern is based on components of the relationships among these components. The structural approach is that it provides a good symbolic description of the image.
- Artificial neural network: connected elements called neurons. This algorithm is non-algorithmic and trainable. The neural network is for pattern classification class is the feed-forward network and others used for classification purposes or conventional neural networks.
- Kernel method: Most import support vector machines, kernel principal component analysis, etc. support vector

machines are a group of supervised learning methods that can be applied to training and testing tests. Different types of kernel functions of SVM are the linear kernel, polynomial kernel, Gaussian radial basis function, and sigmoid.

Classifiers used for character recognition problems nearest neighbor classifier, hidden Markov model (HMM), support vector machine (SVM), artificial neural network (ANN), etc. Jain et al. give a review of statistical pattern recognition techniques. Pal and Singh train a neural network to recognize uppercase handwritten characters based on Fourier descriptors of character boundaries as features. In recognition of handwritten alphabets using neural network and zoning-based diagonal features is addressed.

4. Conclusion

In the 19th century, the growing influence of the “Persian” and “Takri” contribute to the decline of the Sharda Script and Language given Globalisation and language loss, languages revival is said to become more and more relevant in the present century as people seek to recover their cultural autonomy, empower the spiritual and intellectual sovereignty and improve their wellbeing.

Today the Modal to revive Sharda (Indian Manuscript) is not limited to Kashmiri Pandits, who employ the script for ritual purposes and for writing horoscopes but by archeologists/scholars/readers/linguists for

- Preserving this manuscript.
- Train specialists and users to read Sharda for producing critical additions to the texts.
- Research and reach ancient Indian knowledge traditions with emphasis on the broad problem of letter recognition from different manuscripts and adding leftover letters and that too with the use of computer networking

In this paper, we have surveyed the character recognition work that has been done on Devanagari. We have assessed the work done for handwritten characters. We have presented recognition accuracies achieved for character recognition of different scripts. We have seen that the efficient techniques used handwritten characters so that the accuracy of recognition may be increased. Novel features can be proposed to improve the recognition accuracy of different scripts. There is a need to develop the standard database for Devanagari, Sharda Script, etc.

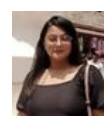
References

- [1] Sir George Grierson, “On the Sharda Alphabet,” Journal of the Royal Asiatic Society, XVII, 1916.
- [2] Jonaraja, “DvitiyaRajatarangini.”
- [3] Al-Biruni, "Tariq-ul-hind," Chronicles of India.
- [4] Abul-Fazl Ibn Mubarak, “”
- [5] Colin Masica, "The Indo-Aryan Languages", Cambridge Language Surveys, 1993.
- [6] Panday Anshuman, “Proposal to encode Sharda script”, UC Berkeley: Department of Linguistics,2009.
- [7] Kalhana,“Rajatarangini”, 1148.
- [8] Dr.VedKumari,“Nilmatha Purana”,2007.
- [9] ShailendraAima “The revival of a lost language”.
- [10] P E Ajmire, Shankar Eknath Warkhede, “An Analytical Study of Devanagari Script Recognition”, 2018.
- [11] Deva (315), Devanagari (Nagari) [online], Available: Devanagari - Wikipedia,
- [12] Ethnologue - Languages of the World [online],Available:http://www.ethnologue.com/show_1language.asp?code=bkc

Author Profile



PURAN BHAT - Originally a Kashmiri Pandit, an inherent domicile of the valley of Kashmir, residing at Kalusa Tehsil and District Bandipur – 193502 in the Union Territory of Jammu and Kashmir- is a student enrolled in Chandigarh University; University Institute of Engineering SAS Nagar Punjab – 140413 pursuing Bachelor of Computer Science Engineering.



Kannagi Rajkhowa- A student of Chandigarh University; University Institute of Engineering SAS Nagar Punjab- 140413 pursuing a Bachelor of Computer Science Engineering is a resident of Nagaon, Assam, 782002. A 20-year-old, who has completed 10th and 12th in 2019 and 2021 respectively from her hometown.