International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Handwritten OCR for Indic Scripts: A Comprehensive Overview of Machine Learning and Deep Learning Techniques

[1]Shaik Moinuddin Ahmed, [2]Abdul Wahid

[1][0000-0001-5395-7648],Department of CSIT,

Maulana Azad National Urdu University, Hyderabad, India

moinuddinahmed@manuu.edu.in

[2][0000-0001-6729-7775], Department of CSIT,

Maulana Azad National Urdu University, Hyderabad, India

awahid@manuu.edu.in

Abstract—The potential uses of cursive optical character recognition, commonly known as OCR, in a number of industries, particularly document digitization, archiving, even language preservation, have attracted a lot of interest lately. In the framework of optical character recognition (OCR), the goal of this research is to provide a thorough understanding of both cutting-edge methods and the unique difficulties presented by Indic scripts. A thorough literature search was conducted in order to conduct this study, during which time relevant publications, conference proceedings, and scientific files were looked for up to the year 2023. As a consequence of the inclusion criteria that were developed to concentrate on studies only addressing Handwritten OCR on Indic scripts, 53 research publications were chosen as the process's outcome. The review provides a thorough analysis of the methodology and approaches employed in the chosen study. Deep neural networks, conventional feature-based methods, machine learning techniques, and hybrid systems have all been investigated as viable answers to the problem of effectively deciphering Indian scripts, because they are famously challenging to write. To operate, these systems require pre-processing techniques, segmentation schemes, and language models. The outcomes of this methodical examination demonstrate that despite the fact that Hand Scanning for Indic script has advanced significantly, room still exists for advancement. Future research could focus on developing trustworthy models that can handle a range of writing styles and enhance accuracy using less-studied Indic scripts. This profession may advance with the creation of collected datasets and defined standards.

Keywords-Handwritten, OCR, Indic scripts, offline recognition, feature extraction.

I. INTRODUCTION

"Optical character recognition" (OCR) refers to the digitization or optical scanning of printed or written text in order to convert it into a machine-readable format. OCR, or optical character recognition, is the short name for a discipline of pattern recognition that is both demanding and interesting. It has a wide range of practical uses. With the development of a retina scanner, Carley created the optical character reader, a primitive method of picture transmission, in 1870. [1]. There are good reasons to be positive about the possibility of improving document identification that draws on printed sources. However, because there is such a wide range of handwriting styles, it can be difficult to identify characters in handwritten writings. It is reasonable to assume that research into the identification of handwritten calligraphy is going to persist in some form in the future as a result. The quality of the method that powers character identification is proportional to the precision of the system used for detection. Character recognition is a complex process that calls for several tools and strategies. OCR, or optical character recognition, has made tremendous progress in recent years, a computationally complex discipline. Huge strides have been made in recent years in both artificial learning and several methods that mainly rely on computation. Most of these developments have been made in this field. The system's objective is to read as quickly and accurately as a human reader while being much more effective. Figure 1 illustrates the several processes that make up the character recognition process. [2].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

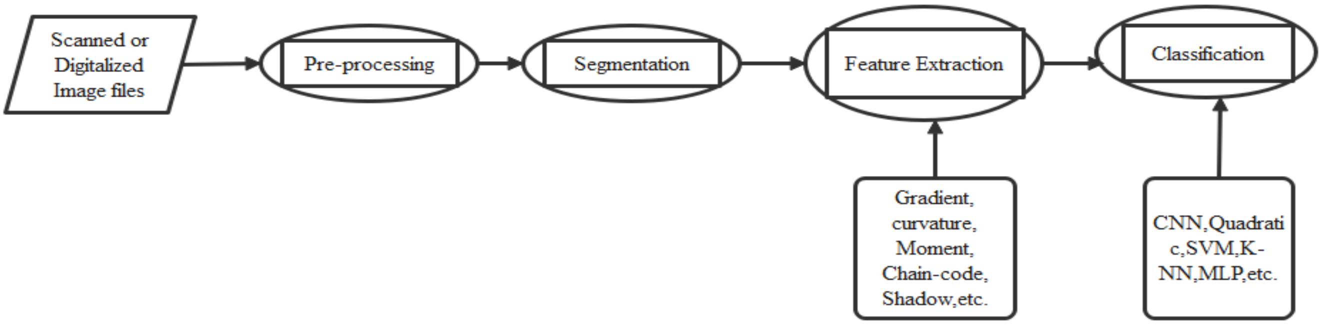

Figure 1: Character recognition procedure stages.

The phases involved in the recognition process are feature extraction, segmentation, the pre-processing and data gathering. Digitising and pre-processing handwritten documents to fix orientation, noise, distortion, slant, skew, binarization, and normalisation is the first step in processing handwritten text. Analytical, or segmentation-based, text processing, and holistic, or segmentation-free, text processing, are the two main types of text processing. The text is divided into lines, words, or ligatures during the segmentation stage. To decrease the amount of data needed for identification and boost recognition strength, the second step, feature extraction, focuses on finding meaningful and unique patterns. The recovered characteristics are what ultimately decide the classifier's accuracy. This is the last step before recognition and categorization, which finds patterns based on incoming data, is the main decision-making stage [3].

A. Background

There are many different languages as well as written systems used in India. The Constitution of India officially recognizes 22 different languages. Dogri, Urdu, Punjabi, Kashmiri, Bengali, Odia, Kannada, Tamil, Nepali, Gujarati, Kannada, Marathi, Bodo, Konkani, Malayalam, Telugu, Sindhi, Assamese, Santali (Santhali), Manipuri, Telugu, and Sindhi are just some of the languages spoken by the people of India. Bodo and Konkani are only two of the many languages spoken in the Indian subcontinent. Gurmukhi, Bengali, Arabic (Perso-Arabic), a Kannada language, Gujarati, Tamil, Malayalam, a dialect of Ass Telugu, and Oriya are only a few examples of the twelve various scripts used to write these languages. Also spoken is English. Scripts can show the articulation and phonetics of a language.

1) Evolution of Indic scripts

The overwhelming majority of Indian applications, including the contemporary ones, have their roots in the ancient Brahmi script. This procedure led to a number of changes. [4], making it Ancient India's earliest writing system. As demonstrated in Fig. 2, these modifications cause new variants of Indian regional letters to develop and diverge.

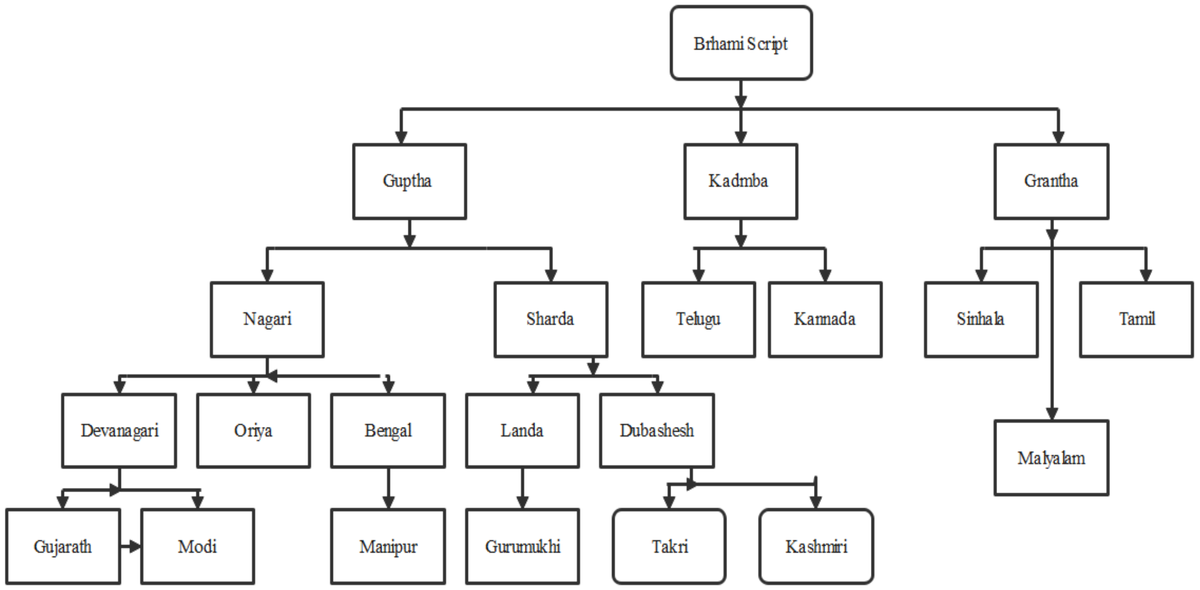

Figure 2: Evolution of Indic scripts[5]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

a) Writing system

Indic scripts are written using either the Alpha syllabary or the Abugida manner of writing. The numerous consonant and vowel pairings that make up this segmental language are what make it up. Consonant letters serve as the foundation and vowel symbols serve as the supporting element in these consonant-vowel combos. A vowel is portrayed with a sound that is different from the others if it's seen by itself or at the beginning of a word.[6].

b) Classification of Indic scripts

One of these classifications comprises the languages that are widely used in north India, and the other category has the scripts who are utilised in south India. The Brahmi script can be divided into these two primary divisions. Multiple languages can be written using the same writing system. Devanagari is used to write languages like Hindi, Dogri, the language of Sanskrit, Kashmiri, Marathi, Sindhi, Nepali, and Konkani, for instance. It also happens to be the Indic alphabet that is most commonly used. [7]. The Bengali alphabet is among the two most extensively utilised in usage worldwide after Devanagari. It is used to write the regional dialects of Assam, Bengali, Manipur, and Maithili.

c) Challenges in Indic scripts

Basic and compound characters dominate Indic scripts. Compound characters are two or more basic characters[8]. Simple letters have simpler forms than compound ones. Gurumukhi and Tamil lack compounds. Figure 3 and 4 show primary Indic script character sets. Most Indian scripts use compound and modified characters. Vowels become diacritics when they precede consonants [9]. Most Indian scripts are left-to-right. Changed characters cannot write left-to-right. Indic scripts lack upper- and lower-case characters. Right-to-left Urdu alphabet. Urdu is Persian Arabic [10]. Although calligraphic, Urdu script lacks upper- and lowercase letters like other Indian scripts. The Persian Arabic script and Urdu script share similarities, but have key differences such as additional letters, diacritics, ligatures, and pronunciation. Arabic OCR systems are not suitable for processing Urdu text due to these differences, so specialized Urdu OCR systems are recommended for accurate text recognition.

At the very top of many Indian inscriptions is a horizontal line called the Matra, also called the Shirorekha [11]. By placing a horizontal line at each letter, it is possible to make a word with a completely linked top bar. The words are inserted without spaces between them to preserve the highest horizontal bar. Indic characters without mantras can be identified from one another. Matra, often referred to as Shirorekha, is a founder of the Bengali script known as Gurumukhi in Devanagari. The languages of Gujarati, Oriya, Kannada, a Telugu, Urdu, Tamil, the language of Malaya and Assamese do not have Shirorekha or Matra characters. Typing in the Matra alphabet may be challenging because OCR strips the characters off the text. Every script utilizing the Indian alphabet has upper, middle, and lower text lines. The main character is in the top zone, which is above the title or shirorekha. The subsidiary characters are in the middle zone, which is below the baseline and below the bottom zone.[12]. In the languages of Gujarati, Tamil, the Oriya language, Kannada, and Malayalam, if there is no headline, the top zone will be above the meanline, and the middle zone will be between the meanline and the baseline. The lowest character on a line of text is represented by the baseline, and the highest character by the meanline. Indic scripts with shirorekha, also known as headlines, are shown in Fig. 5, while Figure 6 shows mean-line-based zones. Table 1 displays similarities as well as distinctions amongst the various Indic scripts.

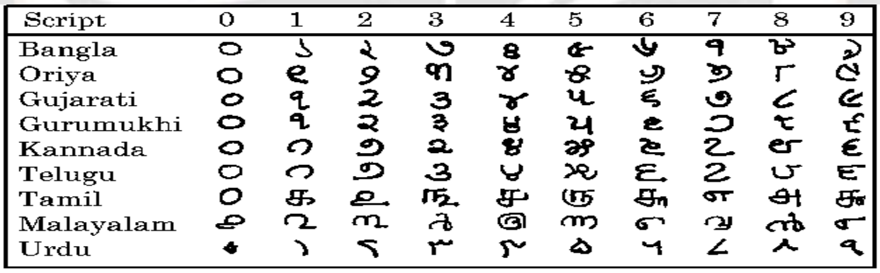

Figure 3: Handwritten vowels from multiple scripts used in India [5].

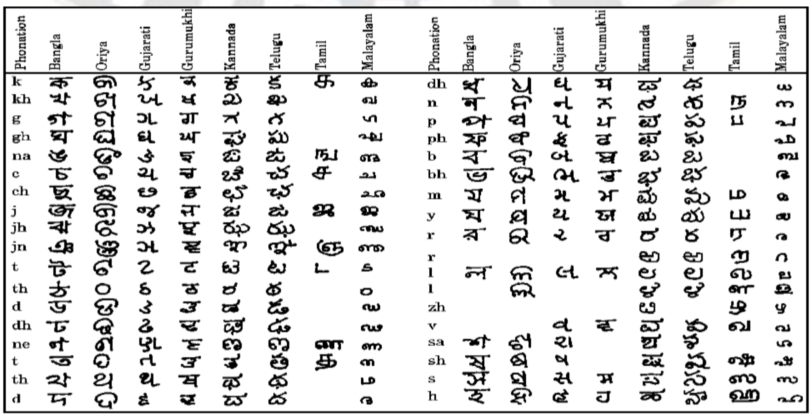

Figure 4 : Written consonants from multiple Indic scripts by hand [13]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

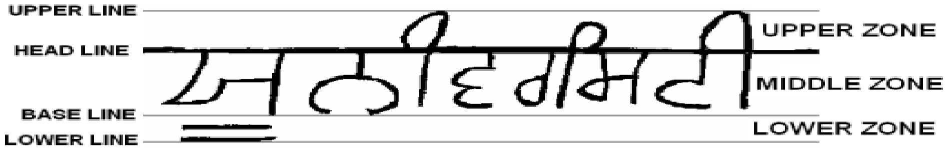

Figure 5: Zones with shirorekha/headline in Gujarati script [14]

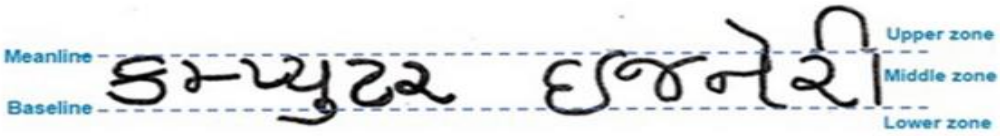

Figure 6: Gujarati zones written in Indic script without headlines or shirorekha [13]

TABLE 1: COMPARISON BETWEEN DEVANAGARI, BANGLA, GURUMUKHI, KANNADA, TELUGU, URDU, ORIYA, TAMIL, MALAYALAM WRITING STYLES.

Script	Origin	Letters	Direction	Features	Common Usage
Devanagari	India	48 letters	Left-to Right	Syllabic, phonetic, with mantras (vowel signs)	Hindi, Sanskrit, Marathi, Nepali
Bangla	India, Bangladesh	44 letters	Left-to Right	Curved, fluid strokes, many ligatures	Bengali, Assamese
Gurumukhi	India, Pakistan	35 letters	Left-to Right	Simple and uniform strokes	Punjabi, Sindhi
Kannada	India	51 letters	Left-to Right	Rounded, flowing strokes	Kannada, Tulu, Kodava
Telugu	India	56 letters	Left-to Right	Syllabic, angular shape	Telugu
Urdu	South Asia	52 letters	Right-to-Left	Connected, Cursive style	Official script of Pakistan
Oriya	India	52 letters	Left-to Right	Distinctive circular shapes, horizontal line connecting letters	Odia
Tamil	India, Sri lanka	247 letters	Left-to Right	Syllabic, with unique vowel representations	Tamil, Malayalam
Malayalam	India	53 letters	Left-to Right	Syllabic, highly curved and circular shapes	Malayalam

The review work offers a comprehensive overview of OCR techniques for Indic scripts, including machine learning and deep learning approaches. It includes recent advancements and a detailed analysis of each approach, emphasizing the importance of handwritten text recognition. The review also explores the role of ‘Deep learning techniques’, particularly ‘Convolutional neural networks’ and ‘Recurrent neural networks’, in revolutionizing OCR for Indic scripts. It also discusses current research trends, emerging methodologies, and potential future developments in OCR for Indic scripts.

The next portions of this paper are organized in a certain sequence. The origins and evolution of important Indic scripts are discussed in the first portion, with an emphasis on the unique challenges and characteristics connected to each. In Section 2, the survey is covered, along with the research goal, questionnaire design, data sources, and search criteria. We have dissected, analyzed, and compared the research into extraction and classification of features in handwritten Indic script recognition in Section 3. Other classification techniques, like those based on machine learning as well as assistance vector machines, are also available. In Section 4, statistical findings of the study is discussed and in section 5 the difficulties of character recognition in Indic code are examined, along with ideas for future study. Section 6 is the paper's final section.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

II. REVIEW METHODS

By defining research questions and selecting relevant studies, the current SLR hopes to locate and summarize literature on OCR. We are going to put the tactics that Kitchenham et al. suggested into action. [15]. Subsequent subsections of the suggested methodology include topics like the review strategy, the standards for inclusion and exclusion, the approach to search, the selection process, etc., as well as techniques for extracting data and synthesising it.

A. *Research questionnaire*

Questionnaires are an integral part of all aspects of research, including articles, projects, surveys, and studies. They provide the study a pointed edge as well as a clear path to go in. The reader will have a much easier time understanding the purpose of this systematic review once the research questions have been formulated. The questionnaire for the research has been meticulously prepared to ensure that it provides complete coverage of all areas of OCR. The following outlines the problems that arise when attempting to OCR handwritten Indic characters.

RQ1: What is handwritten OCR?
RQ2: What are the various methods that are utilized to construct a handwriting OCR template for Indic scripts?
RQ3: When it comes to handwritten Indic script OCR, which feature extraction methods and classification algorithms are the most effective?
RQ4: Scripts used in notable Indian works?

B. *Search criteria*

Searching begins with character recognition-related articles. Year, dataset, feature extraction technique, and classifier group articles. The first pick uses terms like "handwritten character recognition," "Indic script feature extraction," "Indic script classification," etc. We found 1200 articles online. Title, abstract, and citation count are used to further refine the works. This comprehensive investigation included just 53 papers after thorough reading. Fig. 7 shows search criteria for finding relevant publications.

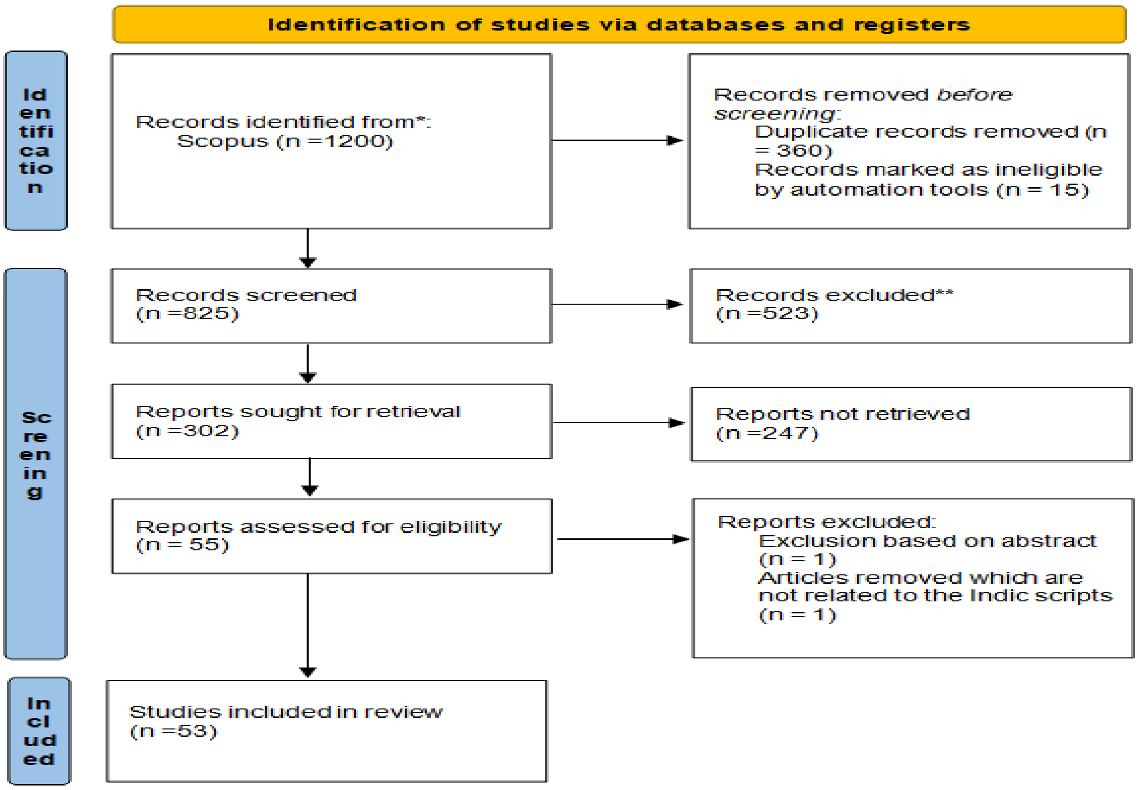

Figure 7: PRISMA flow chart for the selection criteria.

C. *Quality assessment*

The purpose of the quality evaluation step is to identify, through qualitative analysis, which research papers should be retained in a survey and which should be excluded. A complete review, standardized datasets, high-quality research publications, and an emphasis on handwritten Indic script recognition are all necessary for the procedure, It comprises utilizing a quality assessment form to determine the level of quality of selected articles.

D. *Data extraction*

In this stage, 53 chosen studies' metadata was extracted in total. Mendeley and Microsoft Excel were used, as was already noted, to manage this metadata. This phase's main goal was to methodically record the data gathered from the preliminary investigations. The information that was gathered included the research ID, which is used to identify individual studies, study title, authors, publication year, publishing platform (such as journals or conference proceedings), amount of citations, and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

study context, which refers to specific methodologies utilized in the study. The aforementioned data was acquired by means of a thorough investigation of every study, which facilitated the identification of methods and methodologies suggested by the researchers. This procedure also made it easier to categorize the research according to the programming languages that the methods were used with.

III. RESULTS

Based on their commonalities and historical development, the Indic scripts are divided into three groups in this work: scripts from Devanagari, Gurumukhi, and Bengali; scripts from Kannada and Telugu because of their character sets; and scripts from Gujarati, Oriya, Tamil, and Malayalam.

A. Researches conducted on the Devanagari, Gurumukhi, and Bengali scripts

1) Feature extraction

Large input can be transformed into a feature vector using the feature extraction technique, which raises the OCR recognition rate. There are two categories for features: structural and statistical. Depending on the character picture, statistical features including zoning, moments, projection profiles, and histograms are employed for either local or global feature extraction. Characters' topological and geometric structures are connected to structural elements like lines, loops, and intersections. In this essay, we'll go deep into feature extraction techniques for handwritten Devanagari, Gurumukhi, as well as Bangla.One of the first researchers to do the research was Prashant Madhukar Yawalkar. [16] Suggested an updated model for handwritten character recognition that makes use of improved machine learning. Pre-processing, segmentation, feature extraction, and grouping are the building components of the model. Pre-processing, segmentation, as well as feature extraction, are performed on a scanned page written in Devanagari. The Lion Updated GWO, also known as LU-GWO, optimizes weights using a hybrid approach. When compared to employing normal methods, the classification of consonants, numeral, and vowels based on separate characters is superior.

A lot of topological properties were utilised, such as octant centroid features, revised shadow features, including longest run features by Basu et al. [17], a book about handwritten Bengali script that might be among the very first. Two examples of regional gradient descriptors are the ‘Scale Invariant Feature Transform (SIFT)’ as well as the ‘Histogram of Oriented Gradients (HOG)’ for features that have been provided Surinta et al. [18] for recognizing handwritten Bengali characters.

2) Classification and recognition

The three main types of classification techniques are based on networks of neurons (NNs), supported vector neural networks (SVMs), and other approaches. The sections that follow provide an analysis of various strategies.

a) Neural Network based techniques

It is feasible to classify and identify handwritten characters by using methods that rely upon artificial neural networks (ANNs). [19]. Parallel ANN computations are quicker than traditional approaches. These classifiers resemble human cognition. Each neuron in ANN has its own synaptic weights and connections. Input, hidden, and output layers make up these classifiers.

In order to recognize characters manually written in Devanagari script, [20] analyzed pre-trained Deep Convolution Neural Networks (DCNNs) for identifying handwritten Devanagari alphabets using Inception ConvNet, DenseNet, Vgg, and AlexNet. Results show AlexNet outperforms all models with 98% accuracy, while Inception V3 outperforms with 99% accuracy.

To recognize Gurumukhi characters written by hand, [21] recognized using the Hopfield Neural Network (HNN) model. A recognition accuracy of 95.4% has been observed after testing the network sequentially with 1500 distinct input patterns. MATLAB has been used for all preprocessing tasks and Hopfield model implementation.

b) SVM based techniques

Authors of the work Kale et al. [22] Using an SVM-based RBF kernel-based multi-class classifier, handwritten Devanagari characters were recognized. Since basic character precision ratings of 98.50 percent and compound character accuracy scores of 98.51 percent were recorded, this work may be deemed satisfactory.

The study [23] explores offline recognition of Gurmukhi script using feature extraction as well as classification techniques. Principal component analysis (PCA) is used to extract the useful characteristics from the changed division point and peak extent. The recognition accuracy of three SVM kernels—linear, polynomial, and RBF—is investigated. The experiment uses 8960 Gurmukhi character samples from 160 writers and employs five partitioning algorithms and k-fold cross validation approaches. The combination of these classifiers results in a recognition accuracy of 92.3%.[24] provides a gradient and curvature feature-based offline Gurumukhi character recognition method that uses Support Vector Machine to achieve a 98.56% recognition rate. In this study, handwritten Gurumukhi characters are used to train a machine learning system to recognize them using Gabor features. Using a varied kernel size (that is, between 0.64 and 1.28), the authors reported an identification accuracy of 94.29% over 7000 handwritten letter samples. The SIFT algorithm or Gabour filters features were used to train the SVM classifier described in. [25] is a manual character organization system written in the Devanagari script. Using a polynomial SVM classifier and 10 times cross-validation, according to the model, enable it to attain a total recognition accuracy of about 91.39 percent.

3) Miscellaneous techniques

The phrase "miscellaneous techniques" describes any and all techniques that are not included in any of the first two groups. Other strategies include the Hidden Markov Model (also known as HMM) technique, the Quadratic Discriminant Features (QDF) strategy, the kernel-based nearest-neighbor (KNN) methods, the Method of Template Matching (TM) approach, the Mirrored Image Learning (MIL) algorithm and the Decision Tree (DT) predictor. These are but a few of the several strategies that could be applied. A listing of the methodology-related reports is provided below.

a) Quadratic Classifier

[26] proposed HWCR system uses ensembling classifiers to recognize Devanagari characters. The system is comprised of three primary parts: pre-processing, feature extraction, and classification. The ensemble assesses the performance of SVM, K-NN, as well as NN classifiers. The combined output is used to classify the class label. The system achieves an average recognition rate of 88.13% using ensembling. In order to identify complex letters written in Bengali by hand, [27] tests several feature extraction algorithms for handwritten character recognition using the Bangla digit dataset. CAT, HOT, GPB, and BWS techniques are employed, together with a support vector machine for excellent precision. CAT outperforms pixel-based methods in terms of computation time, while the majority voting technique improves performance by achieving 96.8% accuracy.

b) K Nearest Neighbor Classifier

Four handwritten character recognition methods were studied by [28], achieving high accuracy rates of 89.02%, 86.67%, and 95.3%, respectively, contrasted with cutting-edge techniques.

Mirror Image Learning:

To read Devanagari letters written on handwritten paper, [29] utilizing a Mirror Image Training (MIL)-based classifier. The process of establishing a mirror image of a link that is connected to a group of cognitive classes is known as mirror image learning, or MIL. The goal is to expand the other team's data collection so they can learn more from it. A MIL classifier with curvature features was utilised for this work to classify 36,127 handwritten characters with an accuracy of 95.19 percent.

Template Matching:

Template Matching, abbreviated as TM, was used by Bag et al. [30] read Bengali compound characters. This strategy gives composite characters' universal traits. The sample represents the character's shape. This standard template database matched and detected handwritten compound characters. The feature template and standard templates were compared to label the character image. Finally, the highest-scoring standard template labelled the character image. Topological TM recognized 19,800 handwritten characters with 86.74% accuracy.

Table 2 summarizes Devanagari, Gurumukhi, and Bengali script identification findings from classics.

TABLE 2: ACCURATE RECOGNITION OF HANDWRITTEN BENGALI, GURUMUKHI, AND DEVANAGARI CHARACTERS.

Methodology	Data set size	Feature Extraction	Classification Technique	Recognition accuracy (%)
Devanagari characters				
Kale et al. [22]	27,000	Legendre and Zernike moment	Support Vector Machine	98.51(Basic), 98.30(Compound)
Jangid and Srivastava [31]	56,477	Automatic	SL-DCNN	98.00
Pal et al. [32]	36,172	Curvature, Gradient	Mirror Image Learning	95.19(Curvature), 94.94(Gradient)
Gurumukhi characters				
Anupam and Manish [23]	8960	linear-SVM, polynomial-SVM and RBF-SVM	feature extraction and classification techniques	92.3%
Kumar et al. [33]	3500	Power curve fitting	K-Nearest Neighbour	98.10
Bengali basic and compound characters				
Keserwani et al. [34]	41,536	Compound Automatic	Unified-CNN	98.12
Bag et al. [30]	19,800	Compound Topological	Template matching	86.74
Sarkhel et al. [35]	42,697	Compound Multiscale-multicolumn CNN	Support Vector Machine	98.12

B. Research work on Kannada, Telugu and Urdu scripts

1) Feature extraction

The study by Sastry, Panyam Narahari, et al. [36] uses zonal based feature extraction to split character images into predetermined zones and extract statistical features. The characteristics of each zone are used to recognize handwritten characters, and the study uses NNC (nearest neighbor classifier) to identify and classify Telugu text, achieving a recognition accuracy of 78%.

It is advised that the Nastalique language in Urdu be taught using an optical text recognition technology based on assisted learning by Rizvi.S et al., [37]. A number of testing scenarios have examined the proposed system, and those assessments have demonstrated that it can predict training outcomes with an accuracy of 98.4% and test outcomes with an accuracy of 97.3%. This OCR system's detection rate is the highest one ever accomplished for the Urdu language. In particular, the system's OCR software component makes the suggested approach simple to implement. The method that was described can be applied to both typed and handwritten text, this will aid in the creation of future versions of Urdu OCR apps that are more accurate. The CNN method, developed by Najwa et al. [38] utilizing the Arabic Handwriting Characters Set (AHCD), and 88% on the Hijja Dataset, respectively, achieved an accuracy of 97%. We chose Arabic as the language of study because there has been so little previous investigation in this field.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

2) Classification and recognition

a) Techniques based on Neural Network based

[39] The paper offers an Arabic printed OCR system comprised of five stages: pre-processing, feature extraction, character segmentation, classification, as well as post-processing. The system employs a novel chain-code representation mechanism for post-processing, a new thinning algorithm, and a compression-based method called Prediction by Partial Matching (PPM). [40] analyzed handwritten training sets from campus students and the web, dividing each letter into segments. Two techniques were developed for identifying handwritten Kannada characters: The Convolutional Neural Network and the Tesseract software. The results showed 86% accuracy with the Tesseract tool and 87% accuracy with the CNN, with potential improvements based on the chosen data set and additional image processing. It was revealed that a CNN-based template with handwritten Telugu letter recognition had been created. Angadi et al. [41]. The CNN architecture utilised in this study is made up of two fully-connected layers, max-pooling layers and four convolutional layers. Additionally, we employ generalization approaches such as data augmentation and dropout. The SGD optimizer is employed for both model training and validation, alongside the categorical cross- Entropy losses.

Raw pictures and meta-features extracted from the UCOM data were processed and compared during the presentation Asma and Kashif [42] in 2018. In order to train on the Urdu language dataset, both a 'long short-term memory (LSTM)' architecture and a 'convolutional neural network (CNN)' style network based on recurrent neural networks were utilised. According to the study's authors, CNN achieved a precision rate of 97.63% of the thickness graph and 94.82% for the unprocessed images. LSTM offers an accuracy range of 99.33 to 100%, despite this. In 2019, Ahmed et al. [43] ‘One-dimensional recurrent neural network (RNN)’, ‘long-short-term memory (LSTM)’, and ‘bidirectional recurrent neural network (BRNN)’ categorization was designed to distinguish handwritten Nasta'liq Urdu. This technique may identify Nasta'liq handwriting, a variant of Urdu. The researchers also offered a whole new dataset including the writing habits of 500 Urdu-Nasta'liq writers. According to the research, they had been successful in getting very accurate character identification. Between 6.04 and 7.93% of the standard deviation occurred for each of the trials.

b) SVM based techniques

For the objective of decoding Kannada characters written by hand, With their proposal of a support vector machine (SVM), Rajput et al. [44] classifier with five cross-validations that is based on a Gaussian kernel. Chain coding and Fourier descriptor-based normalization of the data was used to feed the multiclass classifier that was employed in this study. To produce the data, the one-versus-rest class notion was used. They asserted that their voice recognition tool had a detection level of 93.92 percent after analyzing 6500 different varieties of handwriting. You can group handwritten Kannada characters into the following categories: In 2012, Pathan et al. [45] relying on the invariant moments method, a solution was developed to identify the handwritten single Urdu letters. It was possible to separate all 36,800 a single- and multi-component letters that were included in the dataset. In the case of multi-component letters, separate invariant moment was determined for the primary and secondary components, respectively. The researchers that used SVM were able to get a total performance rate of 93.59% for the classifications they created.

3) Miscellaneous techniques

a) Quadratic Discriminant Function:

For the purpose of reading scribbled Kannada &Telugu, Pal et al. [29] used a quadratic classification method. In this work, 10,779 Kannada syllables or 10,872 Telugu syllables were recognized by the quadratic predictor using 400-dimensional directional data. These two dialects are both South Indian. Both groups had recognition accuracy rates of at least 90%.

b) K Nearest Neighbor Classifier:

Using a KNN classifier and an evaluation strategy based on Euclidean distance criteria, Sangame et al. [46]were able to classify handwritten Kannada characters. In order to decode the handwritten Kannada letters, Dhandra et al. [47] utilised a classifier with a k-nearest neighbor algorithm and four cross validation rounds. Utilizing spatial data along with a KNN classifier with k = 1 throughout the experimental evaluation led to a recognition reliability of 90.1%. Within the purview of a different investigation, Reefed et al. [48] using a dropout-regularized deep neural network. The recognized ligatures were then grouped together using the K-Means algorithm. According to the paper's authors, their suggested method is much more precise (94.71%) than neural networks in general, which only manage 74.31% accuracy. [49] presents a deep convolutional neural network that can recognize handwritten characters by identifying their low-level textual characteristics. The model achieves 98.7% accuracy, suitable for Android, iOS, RIM, and other languages, and is applicable to OCR systems for all Indian languages.

c) Decision Tree Classifier:

A Decision Tree (DT) classifier trained with 3D features was suggested by Sastry et al. [50] for recognizing handwritten Telugu characters. The SEE5 algorithm was used to create the DT, and it has been rated at 93.10% accuracy when comparing handwritten samples.

A overview of the data from multiple tests utilising classifiers to recognise handwritten Kannada, Telugu, nor Urdu characters is shown in Table 3.

TABLE 3: RECOGNITION ACCURACIES FOR HANDWRITTEN KANNADA, TELUGU AND URDU CHARACTERS

Methodology	**Data set size**	**Feature Extraction**	**Classification Technique**	**Recognition accuracy (%)**
Kannada characters				
Dhandra and Mukarambi [47]	1400	Normalized chain code and wavelet decomposition	K-Nearest Neighbor	95.07%

Fernades and Rodrigues [40]	training sets from campus students and the web	the Tesseract tool	CNN	86% accuracy with the Tesseract tool and 87% accuracy with the CNN
Telugu characters				
Sastry et al. [50]	Not-specified	3D features	Decision Tree	93.10%
Panyam Narahari [36]	Telugu handwritten characters	Zonal based feature Extraction		78%
Angadi et al. [41]	45,133	Automatic	CNN with SGD optimizer	92.40%
Urdu Characters				
Asma and Kashif [42]	Not Specified	Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM)	Neural Network	99.33%
Ahmed et al. [43]	500	recurrent neural networks (RNN)	BLSTM classifier	97.00%
Pathan et al. [45]	36,800	Structural Features	SVM	93.59%
Rafeeq et al. [48]	a dataset containing 17,010 ligatures	Deep neural network	K-Means	94.71%

C. Research work on Gujarati, Oriya, Tamil and Malayalam scripts

1) Feature extraction

Individual picture pixels were handled as characteristics by Prasad et al. [51] in order to recognize handwritten Gujarati letters. Statistical and structural traits have been jointly retrieved by Patel et al. [52]. The structural characteristics were regarded as the major characteristics, whilst the statistical characteristics were regarded as the supporting characteristics. The following elements make up the basic feature set: the total number of picture parts, the quantity of items in the image's top and bottom halves, and the total number of character image holes. Characteristics obtained from the standard deviation, the present, and the centroid distance make up the secondary set of traits.

The creators of the comic were able to create an Oriya character via using Robert's filter on an image of the character [53] in order to determine the gradient's direction and strength. The curved features were then obtained by applying the traditional biquadratic in terpolation procedure. Three levels, other for linear zones, one for curved areas, as well as one for curvy areas, respectively, were quantized for these properties. The produced feature vector's dimension was then decreased using the main element, principal component analysis (PCA), and the data that resulted was then fed into the classifier.

It was suggested that we use the SIFT approach. by Subashini et al. [54] in order to produce a feature vector that is local and invariant from each character image. They used K-means clustering to put together the codebook, using the gathered vector features sets for every letter in the image as their source material. After that, the bag-of-key points approach was used in order to ascertain the overall amount of photographs. In conclusion, SVM serves for the purpose of classifying based on these qualities. From the character boundaries, we were able to generate a total of six statistical features. by Abirami et al. [55] in order to assist character identification. All a free man direction numbers, slant angle, ratio of components, curvature, linearity, and curliness are the six distinguishing characteristics. In a research described in [56], feature selection as well as extraction were performed in two stages. The features were at first chosen using a zoning strategy that used an 8-direction chain-code process.

The following are several methods for Malayalam character recognition: Raju et al. [57] created a technique for feature extraction using gradients and run-counting data (GF-RLC) and three additional fundamental properties: centering position, character code, and aspect ratio. Handwritten Malayalam letters were categorized into the following groups in accordance to, Manjusha et al.[58] developed feature descriptors using a scattering convolution network.

2) Classification and recognition

a) Neural Network based techniques

Rapid learning neural network called the Extreme Learning Machine, or (ELM) [59] a back-propagation-based multi-layer perceptron (BPMLP) was constructed. The BPMLP used features that were based on curvelets to offer training. Based on a dataset made up of 2120 various renditions of handwritten characters, the study's recognition accuracy was found to be 95.99%. In order to make it easier to recognize characters in Malayalam script, Raju et al. [57] a classifier relying on the FFBPNN was suggested. A number of essential characteristics, such as centroid, symbol code, and additional dimension ratio, along with run duration counting the gradient features, were employed in the training process of the sigmoid-activated FFBPNN. Following the analysis of 19,800 manually written characters, FFBPNN was able to reach a 99.78 percent accuracy rate for recognition.

b) SVM based techniques

Several scholars used SVM classifiers to separate handwritten characters in the Indian script. Shanthi et al. [60] Tamil characters may be identified with an accuracy of 82.04% using

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

an SVM classifier that was developed using pixel density characteristics plus max-win voting. Subashini et al. [54] suggested using an SVM classifier to categorize Tamil characters using features from locally invariant SIFT descriptors. [61] By using discrete data values derived from Z-ordering, strip-tree, as well as quad-tree methods, we built a multiclass classification system based on SVM for recognizing Tamil letters. [62] 97.96% of Malayalam letters were recognized using an SVM classifier utilizing a kernel value of 0.02 with 10 rounds conducted cross-validation on 13,200 samples. Manjusha et al. [58] used an SVM classifier in a linear kernel and a scattering multilayer net for identifying 29,302 Malayalam words with 91.05% accuracy.

3) Miscellaneous techniques

a) Quadratic Discriminant Function Classifier:

Wak abayashi et al. [63] It has been shown that it is possible to differentiate between classes using a Quadra Discriminant Function (QDF) classifier with five-fold cross-validation between handwritten Oriya characters. The accuracy of the F-ratio-weighted QDF classifier was 95.14 percent on a sample of 18,190 handwritten characters. Pal et al. [53] The Tamil handwriting was analyzed using a quadratic classifier. This study claims that a quadratic classifier, who was provided with a 400-dimensional linear feature vector to complete the assignment, had a detection rate of 96.73% of 10,216 Tamil letters Moni et al. [64] Malayalam handwriting classification using a modified QDF. When trained on 19,800 hand samples with 12 directional coding variables, the MQDF classifier achieves 95.42 percent accuracy. The MQDF reduces computation costs and increases recognition accuracy by over 10% when compared to the QDF. Raju et al. [57] utilised a streamlined version of QDF to classify Malayalam handwriting. In order to train the SQDF, centroid, character code, & aspect ratio was mixed with run duration count. 99.66% of the 19,800 handwritten characters were recognized by SQDF.

b) K Nearest Neighbor Classifier:

For Gujarati character identification in handwriting, Prasad et al. [65] Here, a weighted KNN predictor was put to use. The suggested method involves adapting the traditional KNN algorithm by using distance metrics like the triangular distance as well as the Euclidean distance with additional feature weights. The proposed classifier recognizes handwritten samples with an accuracy of 86.33 percent using Gabor period XNOR pattern features (GPXNP). BESAC features were used to evaluate the work presented in [66] utilizing the nearest neighbor classifier, was classified. The claimed success rate for the suggested approach in identifying 7800 handwriting Oriya characters was 99.48%.

c) Decision Tree Classifier:

[67] artificial neural networks are used to recognise both typed and handwritten Gujarati conjunct characters. The study achieves a success rate of 99.4% and 94.1% in classifying these characters.

Classifier results for handwriting Gujarati, Oriya, Tamil, etc. Malayalam characters are outlined in Table 4.

TABLE 4: RECOGNITION ACCURACIES FOR HANDWRITTEN GUJARATI, ORIYA, TAMIL AND MALAYALAM SCRIPTS

Methodology	Data set size	Feature Extraction	Classification Technique	Recognition accuracy (%)
Gujarati characters				
Patel and Desai [52]	Unspecified	Centroid and moment based features	Tree classifier and KNN	63.10
Prasad and Kulkarni [51]	16,560	GPXNP	Adaptive NFC using feature selection	68.67
Oriya characters				
Pal et al. [53]	18,190	Gradient, Curvature and PCA	Quadratic classifier	94.60
Tamil Characters				
Subashini and Kodikara [54]	8000	SIFT feature descriptors	SVM	81.62
Abirami et al. [55]	3360	Freeman directional code, curvature, etc.	HMM	85.00
Malayalam Characters				
Chacko et al.[62]	9000	Wavelet Energy features	Extreme Learning Machine	95.59
John et al. [68]	13,200	Gradient, Curvature and PCA	SVM with RBF kernel	97.96
Manjusha et al. [58]	29,302	Linear Kennel	SVM	91.05%

IV. STATISTICAL ANALYSIS OF THE FINDINGS

Devanagari, Gurumukhi, Bengali, Kannada, Telugu, Urdu, Gujarati, Oriya, Tamil, and Malayalam are only few of the scripts whose characters are examined in depth throughout the review. Since these scripts are so popular in India, effective character recognition is crucial for a wide range of uses.

Character recognition relies heavily on a process known as feature extraction. Researchers have used a wide variety of methods to attempt to capture the unique qualities of handwritten characters in these scripts. Legendre and Zernike moment extraction, directional, LBP, and regional features, curvature and gradient analysis, as well as SIFT feature descriptors are all common methods. These strategies are designed to help classify photographs of characters by extracting useful information from them.

Recognizing handwritten characters successfully has been accomplished through the use of several categorization strategies. Neural network techniques, support vector machine methods, and other approaches fall into one of these three broad categories.

A. Applications of Neural Networks

- To improve recognition of Devanagari and Gurumukhi characters, researchers have turned to deep convolution neural networks (DCNN).
- The recognition of Urdu characters by Extreme Learning Machines (ELM) has shown remarkable accuracy.
- Both the Tamil and Malayalam scripts have been used with BPNN and MLP (Multilayer Perceptron's) for character recognition.

Character recognition using SVMs trained with a variety of kernels has proven successful for languages written in scripts as diverse as Devanagari, Gurumukhi, Telugu, Kannada, and Malayalam. Features such as pixel density, SIFT descriptors, and structural features have been employed in conjunction with SVM classifiers. Classification using the Quadratic Discriminant Function (QDF) has shown promising results when applied to the Oriya, Tamil, and Malayalam scripts. Classifiers based on K Nearest Neighbors (KNN) have been used for recognizing characters in Kannada, Telugu, and Oriya. Bengali and Gujarati scripts have both benefited from template matching methods. The Gujarati script has been classified using Decision Trees.

Accuracy of Recognition:

The precision of different scripts and approaches to recognition varies widely. Optical text recognition systems have reached an accuracy of 98.4 percent while processing Urdu text, and SVM classifiers have achieved 98.51% when processing Devanagari text. Recognition rates greater than 99% have been achieved for both Tamil and Malayalam scripts. Accuracy rates of 99.48% and 94.6% have been achieved for the Gujarati and Oriya scripts, respectively.

B. Challenges and Future Perspectives

- The report shows how academic scholars have used feature extraction and classification to decipher handwritten Indic characters. It's great that recognition accuracy is high, but there are still difficulties to overcome, such as limited databases, literature written in multiple scripts, strange characters, and more. Character recognition using the Indic script needs to be improved in the following areas:
- Standardised datasets are needed to evaluate algorithms. Research will benefit from accessible and reliable datasets.
- Addressing confusing and similar characters: Some Indian script characters seem alike, making identification difficult. A two-step recognition algorithm groups related characters and classifies them by property to increase accuracy.
- To conserve and digitise significant manuscript collections, further research is required to recognise deteriorating, noisy, and historical materials.
- Future study should investigate optimum ensembles for classification performance.
- Due to structural complexity, Indic character recognition is error-prone. Grapheme properties, linguistic data, and script-specific information may improve accuracy, and post-recognition mistake detection and correction should be prioritised.
- Commercial tools may provide adequate performance for standard fonts and scripts, but their accuracy may decrease significantly when applied to complex handwritten Indic scripts.

V. Conclusion

Within the purview of this study, we have conducted a comprehensive examination of different methods for character extraction and classification strategies used in the identification of handwritten Indic scripts. Our results highlight a critical realization: The accuracy of model recognition is directly impacted by the type and volume of data gathered. One of the primary issues that arises when working with Indic scripts is the lack of benchmark datasets. This discovery highlights a critical component of this field: the urgent need to create extensive, well-managed databases that are specifically designed for Indic scripts. Our investigations have thoroughly examined the efficacy of various recognition techniques. Examining and evaluating a wide range of feature extraction and classification methods allowed for this assessment. The recognition rates of each of these strategies served as the primary evaluation metric. As a result of this thorough analysis, We have also found a number of issues that arise from the use of Indic scripts, which paves the way for further research targeted at coming up with creative fixes for these problems. Given the extensive research conducted for this paper, it is clear that using a hybrid approach to feature extraction and classification techniques is essential to getting the most accurate results possible. When we refer to "hybrid feature extraction," we mean the blending of multiple methods that separately extract and examine different data elements. Because of the intrinsic complexity and diversity of Indic scripts, a hybrid approach is necessary. These scripts frequently show nuanced correlations between variables, the kind of interactions that are easier to identify and comprehend when several extraction and classification techniques are used.

References

[1] J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," *IEEE Access,* vol. 8, pp. 142642–142668, 2020, doi: 10.1109/ACCESS.2020.3012542.

[2] T. P. Singh, S. Gupta, and M. Garg, "A Review on Online and Offline Handwritten Gurmukhi Character Recognition," in

2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2022, pp. 1–6. doi: 10.1109/ICRITO56286.2022.9964657.

[3] Z. Xie *et al.*, "Weakly supervised precise segmentation for historical document images," *Neurocomputing*, vol. 350, pp. 271–281, 2019, doi: https://doi.org/10.1016/j.neucom.2019.04.001.

[4] R. Sharma, B. N. Kaushik, and N. K. Gondhi, "Devanagari and gurmukhi script recognition in the context of machine learning classifiers," *J. Artif. Intell.*, vol. 11, no. 2, pp. 65–70, 2018, doi: 10.3923/jai.2018.65.70.

[5] U. Pal, R. Jayadevan, and N. Sharma, "Handwriting Recognition in Indian Regional Scripts: A Survey of Offline Techniques," *ACM Trans. Asian Lang. Inf. Process.*, vol. 11, no. 1, Mar. 2012, doi: 10.1145/2090176.2090177.

[6] S. Dargan and M. Kumar, "Writer Identification System for Indic and Non-Indic Scripts: State-of-the-Art Survey," *Arch. Comput. Methods Eng.*, vol. 26, no. 4, pp. 1283–1311, 2019, doi: 10.1007/s11831-018-9278-z.

[7] M. Yadav, R. K. Purwar, and M. Mittal, "Handwritten Hindi character recognition: A review," *IET Image Process.*, vol. 12, no. 11, pp. 1919–1933, 2018, doi: 10.1049/iet-ipr.2017.0184.

[8] S. M. Obaidullah, K. C. Santosh, C. Halder, N. Das, and K. Roy, "Automatic Indic script identification from handwritten documents: page, block, line and word-level approach," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 1, pp. 87–106, 2019, doi: 10.1007/s13042-017-0702-8.

[9] A. K. Sharma, D. M. Adhyaru, and T. H. Zaveri, "A Survey on Devanagari Character Recognition BT - Smart Systems and IoT: Innovations in Computing," 2020, pp. 429–437.

[10] N. H. Khan and A. Adnan, "Urdu Optical Character Recognition Systems: Present Contributions and Future Directions," *IEEE Access*, vol. 6, pp. 46019–46046, 2018, doi: 10.1109/ACCESS.2018.2865532.

[11] N. R. Soora and P. S. Deshpande, "A novel local skew correction and segmentation approach for printed multilingual Indian documents," *Alexandria Eng. J.*, vol. 57, no. 3, pp. 1609–1618, 2018, doi: https://doi.org/10.1016/j.aej.2017.06.010.

[12] P. P. Roy, A. K. Bhunia, A. Das, P. Dey, and U. Pal, "HMM-based Indic handwritten word recognition using zone segmentation," *Pattern Recognit.*, vol. 60, pp. 1057–1075, 2016, doi: https://doi.org/10.1016/j.patcog.2016.04.012.

[13] R. Sharma and B. Kaushik, "Offline recognition of handwritten Indic scripts: A state-of-the-art survey and future perspectives," *Comput. Sci. Rev.*, vol. 38, p. 100302, 2020, doi: 10.1016/j.cosrev.2020.100302.

[14] P. Jindal and B. Jindal, "Line and Word Segmentation of handwritten text documents written in Gurmukhi Script using mid point detection technique," *2015 2nd Int. Conf. Recent Adv. Eng. Comput. Sci. RAECS 2015*, pp. 11–19, 2016, doi: 10.1109/RAECS.2015.7453388.

[15] B. Kitchenham *et al.*, "Systematic literature reviews in software engineering-A tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010, doi: 10.1016/j.infsof.2010.03.006.

[16] P. M. Yawalkar and M. U. Kharat, "Automatic handwritten character recognition of Devanagari language: a hybrid training algorithm for neural network," *Evol. Intell.*, vol. 15, no. 2, pp. 1499–1516, 2022, doi: 10.1007/s12065-021-00597-8.

[17] S. Basu, N. Das, R. Sarkar, M. Kundu, M. Nasipuri, and D. K. Basu, "A hierarchical approach to recognition of handwritten Bangla characters," *Pattern Recognit.*, vol. 42, no. 7, pp. 1467–1484, 2009, doi: https://doi.org/10.1016/j.patcog.2009.01.008.

[18] O. Surinta, M. F. Karaaba, L. R. B. Schomaker, and M. A. Wiering, "Recognition of handwritten characters using local gradient feature descriptors," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 405–414, 2015, doi: https://doi.org/10.1016/j.engappai.2015.07.017.

[19] M. Labani, P. Moradi, F. Ahmadizar, and M. Jalili, "A novel multivariate filter method for feature selection in text classification problems," *Eng. Appl. Artif. Intell.*, vol. 70, pp. 25–37, 2018, doi: https://doi.org/10.1016/j.engappai.2017.12.014.

[20] N. Aneja and S. Aneja, "Transfer Learning using CNN for Handwritten Devanagari Character Recognition," *1st IEEE Int. Conf. Adv. Inf. Technol. ICAIT 2019 - Proc.*, pp. 293–296, 2019, doi: 10.1109/ICAIT47043.2019.8987286.

[21] P. K. Sarangi, A. K. Sahoo, S. R. Nayak, A. Agarwal, and A. Sethy, "Recognition of Isolated Handwritten Gurumukhi Numerals Using Hopfield Neural Network BT - Computational Intelligence in Pattern Recognition," 2022, pp. 597–605.

[22] K.V.Kale, S.V.Chavan, M.M.Kazi, and Y.S.Rode, "Handwritten and Printed Devanagari Compound using Multiclass SVM Classifier with Orthogonal moment Feature," *Int. J. Comput. Appl.*, vol. 71, no. 24, pp. 31–37, 2013.

[23] A. Garg, M. K. Jindal, and A. Singh, "Offline handwritten Gurmukhi character recognition: k-NN vs. SVM classifier," *Int. J. Inf. Technol.*, vol. 13, no. 6, pp. 2389–2396, 2021, doi: 10.1007/s41870-019-00398-4.

[24] A. Aggarwal and K. Singh, "Handwritten Gurmukhi character recognition," *IEEE Int. Conf. Comput. Commun. Control. IC4 2015*, no. February, 2016, doi: 10.1109/IC4.2015.7375678.

[25] S. R. Narang, M. K. Jindal, S. Ahuja, and M. Kumar, "On the recognition of Devanagari ancient handwritten characters using SIFT and Gabor features," *Soft Comput.*, vol. 24, no. 22, pp. 17279–17289, 2020, doi: 10.1007/s00500-020-05018-z.

[26] A. Deore, S. P.; Pravin, "Ensembling: Model of histogram of oriented gradient based handwritten devanagari character recognition system," *Trait. du Signal*, vol. 34, no. 1/2, pp. 7–20, 2017.

[27] O. Surinta, L. Schomaker, and M. Wiering, "A comparison of feature and pixel-based methods for recognizing handwritten bangla digits," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2013, pp. 165–169. doi: 10.1109/ICDAR.2013.40.

[28] D. Mahapatra, C. Choudhury, and R. K. Karsh, "Handwritten Character Recognition Using KNN and SVM Based Classifier over Feature Vector from Autoencoder," *Commun. Comput. Inf. Sci.*, vol. 1240 CCIS, no. February 2023, pp. 304–317, 2020, doi: 10.1007/978-981-15-6315-7_25.

[29] U. Pal, T. Wakabayashi, and F. Kimura, "Comparative Study of Devnagari Handwritten Character Recognition Using Different Feature and Classifiers," in *2009 10th International Conference on Document Analysis and Recognition*, 2009, pp. 1111–1115. doi: 10.1109/ICDAR.2009.244.

[30] S. Bag, G. Harit, and P. Bhowmick, "Recognition of Bangla compound characters using structural decomposition," *Pattern Recognit.*, vol. 47, no. 3, pp. 1187–1201, 2014, doi: https://doi.org/10.1016/j.patcog.2013.08.026.

[31] M. Jangid and S. Srivastava, "Handwritten Devanagari character recognition using layer-wise training of deep convolutional neural networks and adaptive gradient methods," *J. Imaging*, vol. 4, no. 2, 2018, doi:

10.3390/jimaging4020041.

[32] N. Sharma, U. Pal, F. Kimura, and S. Pal, "Recognition of Off-Line Handwritten Devnagari Characters Using Quadratic Classifier BT - Computer Vision, Graphics and Image Processing," 2006, pp. 805–816.

[33] M. Kumar, R. K. Sharma, and M. K. Jindal, "Efficient Feature Extraction Techniques for Offline Handwritten Gurmukhi Character Recognition," *Natl. Acad. Sci. Lett.*, vol. 37, no. 4, pp. 381–391, 2014, doi: 10.1007/s40009-014-0253-4.

[34] P. Keserwani, T. Ali, and P. P. Roy, "Handwritten Bangla character and numeral recognition using convolutional neural network for low-memory GPU," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 12, pp. 3485–3497, 2019, doi: 10.1007/s13042-019-00938-1.

[35] R. Sarkhel, N. Das, A. K. Saha, and M. Nasipuri, "A multi-objective approach towards cost effective isolated handwritten Bangla character and digit recognition," *Pattern Recognit.*, vol. 58, pp. 172–189, 2016, doi: https://doi.org/10.1016/j.patcog.2016.04.010.

[36] P. N. Sastry, T. R. V Lakshmi, N. V. K. Rao, T. V Rajinikanth, and A. Wahab, "Telugu Handwritten Character Recognition Using Zoning Features," in *2014 International Conference on IT Convergence and Security (ICITCS)*, 2014, pp. 1–4. doi: 10.1109/ICITCS.2014.7021817.

[37] S. S. R. Rizvi, A. Sagheer, K. Adnan, and A. Muhammad, "Optical Character Recognition System for Nastalique Urdu-Like Script Languages Using Supervised Learning," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 10, 2019, doi: 10.1142/S0218001419530045.

[38] M. Agrawal, B. Chauhan, and T. Agrawal, "Machine Learning Algorithms for Handwritten Devanagari Character Recognition: A Systematic Review," *J. Sci. Technol.*, vol. 7, no. 01, pp. 2456–5660, 2022, [Online]. Available: www.jst.org.indoi:https://doi.org/10.46243/jst.2022.v7.i01.pp01-16

[39] M. Alghamdi, "A Novel Approach to Printed Arabic Optical Character Recognition Alghamdi , Mansoor Award date : DOCTOR OF PHILOSOPHY A Novel Approach to Printed Arabic Optical Character Recognition Alghamdi , Mansoor Award date :," *Arab. J. Sci. Eng.*, vol. 47, no. 2, pp. 2219–2237, 2022.

[40] R. Fernandes and A. P. Rodrigues, "Kannada Handwritten Script Recognition using Machine Learning Techniques," in *2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, 2019, pp. 1–6. doi: 10.1109/DISCOVER47552.2019.9008097.

[41] A. Angadi, V. Kumari, and S. Keerthi, "A Deep Learning Approach to Recognize Handwritten Telugu Character Using Convolution Neural Networks," *Proc. 4th Int. Conf. Comput. Manag.*, no. Iccm, pp. 8–12, 2018.

[42] A. Naseer and K. Zafar, "Comparative analysis of raw images and meta feature based Urdu OCR using CNN and LSTM," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 1, pp. 419–424, 2018, doi: 10.14569/IJACSA.2018.090157.

[43] S. Bin Ahmed, S. Naz, S. Swati, and M. I. Razzak, "Handwritten Urdu character recognition using one-dimensional BLSTM classifier," *Neural Comput. Appl.*, vol. 31, no. 4, pp. 1143–1151, 2019, doi: 10.1007/s00521-017-3146-x.

[44] G. G. Rajput and R. Horakeri, "Shape descriptors based handwritten character recognition engine with application to Kannada characters," in *2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011)*, 2011, pp. 135–141. doi: 10.1109/ICCCT.2011.6075175.

[45] R. J. Pathan, I. K., Ali, A. A., & Ramteke, "Recognition of offline handwritten isolated Urdu character," *Adv. Comput. Res.*, vol. 4, no. 1, 2012.

[46] S. K. Sangame, R. J. Ramteke, and B. Rajkumar, "Recognition of isolated handwritten Kannada vowels," *Adv. Comput. Res.*, Jan. 2009.

[47] G. Mukarambi, B. V Dhandra, and M. Hangarge, "A Zone Based Character Recognition Engine for Kannada and English Scripts," *Procedia Eng.*, vol. 38, pp. 3292–3299, 2012, doi: https://doi.org/10.1016/j.proeng.2012.06.381.

[48] M. J. Rafeeq, Z. ur Rehman, A. Khan, I. A. Khan, and W. Jadoon, "Ligature categorization based Nastaliq Urdu recognition using deep neural networks," *Comput. Math. Organ. Theory*, vol. 25, no. 2, pp. 184–195, 2019, doi: 10.1007/s10588-018-9271-y.

[49] T. M. Kumari and A. V. Babu, "Recognition of offline hand written telugu script using deep learning," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 3307, pp. 2456–3307, 2021.

[50] P. N. Sastry, R. Krishnan, B. Venkata, and S. Ram, "Classification and Identification of Telugu Handwritten Characters Extracted From Palm Leaves Using Decision Tree Approach," *Citeseer*, vol. 5, no. 3, 2010, [Online]. Available: www.arpnjournals.com

[51] J. R. Prasad, U. V Kulkarni, and R. S. Prasad, "Template Matching Algorithm for Gujrati Character Recognition," in *2009 Second International Conference on Emerging Trends in Engineering & Technology*, 2009, pp. 263–268. doi: 10.1109/ICETET.2009.220.

[52] S. B. Patel, T. Ghosh, A. Dutta, and S. Singh, "Stress analysis in 3D IC having Thermal Through Silicon Vias (TTSV)," in *2013 IEEE 63rd Electronic Components and Technology Conference*, 2013, pp. 2337–2341. doi: 10.1109/ECTC.2013.6575910.

[53] U. Pal, T. Wakabayashi, and F. Kimura, "A System for Off-Line Oriya Handwritten Character Recognition Using Curvature Feature," 2007, pp. 227–229. doi: 10.1109/icit.2007.63.

[54] A. Subashini and N. D. Kodikara, "Bag-of-keypoints approach for Tamil handwritten character recognition using SVMs," in *2011 International Conference on Advances in ICT for Emerging Regions (ICTer)*, 2011, pp. 102–107. doi: 10.1109/ICTer.2011.6075033.

[55] S. Abirami, V. Essakiammal, and R. Baskaran, "Statistical features based character recognition for offline handwritten Tamil document images using HMM," *Int. J. Comput. Vis. Robot.*, vol. 5, no. 4, pp. 422–440, 2015, doi: 10.1504/IJCVR.2015.072192.

[56] S. M. Shyni, M. Antony Robert Raj, and S. Abirami, "Offline tamil handwritten character recognition using sub line direction and bounding box techniques," *Indian J. Sci. Technol.*, vol. 8, no. April, pp. 110–116, 2015, doi: 10.17485/ijst/2015/v8iS7/67780.

[57] G. RAJU, B. S. MONI, and M. S. NAIR, "A novel handwritten character recognition system using gradient based features and run length count," *Sadhana*, vol. 39, no. 6, pp. 1333–1355, 2014, doi: 10.1007/s12046-014-0274-1.

[58] K. Manjusha, M. A. Kumar, and K. P. Soman, "On developing handwritten character image database for Malayalam language script," *Eng. Sci. Technol. an Int. J.*, vol. 22, no. 2, pp. 637–645, 2019, doi: https://doi.org/10.1016/j.jestch.2018.10.011.

[59] S. M. Salaken, A. Khosravi, T. Nguyen, and S. Nahavandi, "Extreme learning machine based transfer learning

algorithms: A survey," *Neurocomputing*, vol. 267, pp. 516–524, 2017, doi: https://doi.org/10.1016/j.neucom.2017.06.037.

[60] N. Shanthi and K. Duraiswamy, "A novel SVM-based handwritten Tamil character recognition system," *Pattern Anal. Appl.*, vol. 13, no. 2, pp. 173–180, 2010, doi: 10.1007/s10044-009-0147-0.

[61] M. A. R. Raj and S. Abirami, "Structural representation-based off-line Tamil handwritten character recognition," *Soft Comput.*, vol. 24, no. 2, pp. 1447–1472, 2020, doi: 10.1007/s00500-019-03978-5.

[62] B. P. Chacko, V. R. Vimal Krishnan, G. Raju, and P. Babu Anto, "Handwritten character recognition using wavelet energy and extreme learning machine," *Int. J. Mach. Learn. Cybern.*, vol. 3, no. 2, pp. 149–161, 2012, doi: 10.1007/s13042-011-0049-5.

[63] T. Wakabayashi, U. Pal, F. Kimura, and Y. Miyake, "F-ratio Based Weighted Feature Extraction for Similar Shape Character Recognition," in *2009 10th International Conference on Document Analysis and Recognition*, 2009, pp. 196–200. doi: 10.1109/ICDAR.2009.197.

[64] B. S. Moni and G. Raju, "Modified Quadratic Classifier and Directional Features for Handwritten Malayalam Character Recognition," *IJCA Spec. Issue "Computational Sci. - New Dimens. Perspect.*, pp. 30–34, 2011.

[65] J. R. Prasad and U. Kulkarni, "Gujrati character recognition using weighted k-NN and Mean $\chi 2$ distance measure," *Int. J. Mach. Learn. Cybern.*, vol. 6, no. 1, pp. 69–82, 2015, doi: 10.1007/s13042-013-0187-z.

[66] K. S. Dash, N. B. Puhan, and G. Panda, "BESAC: Binary External Symmetry Axis Constellation for unconstrained handwritten character recognition," *Pattern Recognit. Lett.*, vol. 83, pp. 413–422, 2016, doi: https://doi.org/10.1016/j.patrec.2016.05.031.

[67] B. C. Patel, "Identification of typewritten and handwritten Conjunct Gujarati characters using artificial neural network," *Int. J. Appl. Pattern Recognit.*, vol. 7, no. 1, pp. 24–40, Jan. 2022, doi: 10.1504/IJAPR.2022.122267.

[68] J. John, K. Balakrishnan, and P. K. V, "A System for Offline Recognition of Handwritten Characters in Malayalam Script," *Int. J. Image, Graph. Signal Process.*, vol. 5, no. 4, pp. 53–59, 2013, doi: 10.5815/ijigsp.2013.04.07.