A Comparative Study on Brahmi Script Character Recognition

Nomitha Chawla
Computer Science and Engineering Department,
Bapuji Institute of Engineering and Technology,
Davangere,India
nomitachawla@gmail.com

Nirmala C R
Computer Science and Engineering Department,
Bapuji Institute of Engineering and Technology,
Davangere,India
nirmala.cr@gmail.com

Abstract **-**The Brahmi script is a crucial aspect of our cultural heritage, being an important historical and linguistic influence on contemporary Indian scripts. This review paper aims to focus on how the implementation of Artificial Intelligence in the form of an Optical Character Recognition (OCR) system has been developed to decode inscriptions in an ancient language. This process is done by extracting and identifying Brahmi characters. These inscriptions pose additional challenges, with a wide variety of forms of the script used, complex ligatures, and historical variation, combined with problems such as the effects of erosion, weathering, and changes in the directions and intensities of incoming light. In this paper various studies and approaches were investigated including image processing, Machine Learning (ML) and Deep Learning (DL) for character recognition. The analysis of present-day methods summarizes the strong points, drawbacks, and research requirements with a focus on the essential challenges in improving the image-preparation, segmentation, character recognition, and language-computing using AI-aided approach. These findings are significant not only for the purposes of preserving Brahmi script inscriptions through digitization because it makes our ancient scripts in readable form, but also for the advancement of historical document studies in general and broad scholarly engagement and interpretation.

Keywords-Brahmi Script; Artificial Intelligence; Machine Learning; Deep Learning; Character Extraction, Character Recognition.

I. INTRODUCTION

The Brahmi script is widely regarded as one of the oldest scripts in the world. This has been widely practiced from ancient India and is well known for the same. Brahmi is considered the parent of many modern Indic scripts, due to its long-lasting influence on the development of written languages. It has roots–literally–centuries old [2], [4],[5]. Brahmi characters feature detailed strokes and ligatures that require careful craftsmanship and precision, highlighting the value placed on written language in ancient societies and the artistry associated with their writing systems.

In recent years, there has been a verified growth and interest in the preservation and digital investigation of those old scripts such as Brahmi which themselves have been forming to perpetuate in recent time. As more and more digital images of ancient books and stone inscriptions become available, we need means with which to get characters out of the images and recognized. This would require knowing everything there is known about the structure, design, and variations of the script. The digitization process of ancient Indian scripts, as part of the government of India's language documentation and digital archiving program becomes a critical aspect preserving this 'ethos' of linguistic as well as historical documents that are instrumental for the future generation [6].

Brahmi script is an important ancient character set used in several South Asian languages, but extraction, recognition, and understanding scripts like Brahmi script has not received as much attention as that given to Devanagari or Bengali. This is partly because the script is highly complex, with an unintuitive amalgamation of consonants, vowels, and etchings [8].

Brahmi inscriptions on stone, a fascinating treasure of cultural-spiritual, linguistic and historical relevance. These inscriptions are important artifacts that help unravel the mysteries of the past. These inscriptions witness the glimpse of historical events and culture practices as well as religious beliefs, administrative systems and artistic expression in antiquity. The exact preservation and interpretation of these inscriptions are an important part in reconstructing the tapestry of ancient Indian societies and providing insights into their intellectual, spiritual, and material endeavors [12].

Recent developments in machine learning and computer vision have made it feasible to build into autonomous systems the capability to extract and identify characters in images. This paper provides a comprehensive review of the recent developments in character extraction and recognition techniques for Brahmi script, and it aims to address the lack of studies on this subject. This research focuses on technologies developed for detecting and analyzing inscriptions of Brahmi script on stone. This

technology is known as Optical Character Recognition (OCR). It is an interesting step forward and the use of OCR systems for Brahmi script inscriptions will only be a historic leap in unlocking the knowledge hidden in these ancient artifacts as innovation continues its journey to growth. AI-driven OCR technologies enable researchers and scholars to unlock the complex details embedded in Brahmi script inscriptions, revealing the messages and narratives encoded within with newfound precision and efficiency. The combination of these many disciplines is giving promise to the study of Brahmi, and as an interdisciplinary area where linguistics, archaeology, AI and other cultural heritage preservation methods work together opens many new frontiers in the exploration of the multiple dimensions of this ancient script and its lasting influence among ancient scripts and languages [3],[5].

Overall, the primary objective of the present review paper is to analyze the current existing methodology that is employed in the character extraction and recognition methods of Brahmi script. By identifying gaps in research and opportunities for future study, the paper aims to advance the ways in which we protect and analyze such an essential component of our cultural heritage. An aid for academic research, digitization as well as preservation of Brahmi inscribed stones, based on OCR systems. The review outlines the current status, methodologies applied, and experiences gained thus far in this journey towards accurate reading and understanding of the ancient Brahmi inscriptions.

II. Analysis of Methods Used

Vincen et al. introduce a research report on the challenges faced by many Indonesians in learning to read and write the Brahmi script. [1]. They recommend a software program for the recognition of scripts. Java and Bali: The characters of Java and Bali are derived from Brahmi outline used in ancient India. The attempts in scripting classification produced numerical outputs and needed letter classification.

This paper employs a Convolutional Neural Network (CNN) for the classification of the Brahmi script. For this, they are using a deep neural network model, VGG16 architecture. Some of the key steps within the approach include:

1. Pre-processing the input data by zero padding, scaling up to 224 x 224 - grayscale - binarization.
2. Data augmentation can be used for replacing the data and therefore reduce the bias of model.
3. The VGG16 model is then trained on labeled data. Next step is to split the dataset into 2 parts: 70% for training and 30% for testing. The model is trained with some hyper parameters, as learning rate of 0.00001, batch size 25 and training for 10 epochs.
4. A confusion matrix approach is used to assess performance in terms of accuracy, recall, and precision.

The classification assessments produce good results, with the highest accuracy rate of 96 % and both the recall and precision rate of 98 %. It mentioned that increasing the size of training data can improve predictive performance. In the future, they can increase datasets or rather focus on sentence recognition and give already written sentence prompts back based on the learnt vocabulary in written phrases and integrating auditory feedback for script spelling [1].

In an exploration paper, Aniket et al. [2] emphasize the difficulties associated with the identification and segmenting of touching Brahmi characters obtained from the degraded digital images and proposed an algorithm for OCR using ensemble classification technique.

For the purpose of testing, the dataset used in this study was a total of one hundred and fourteen characters collected from Archaeological Survey of India. These were touching and non-touching characters. A vital starting point for recognizing ancient Brahmi characters was to put in place a methodology. Researchers used a mix of hand-curated data in the beginning, creating a dataset of 287 characters with 210 examples for each character for a total of 60,270 images. Several preprocessing techniques including noise elimination, grayscale transformation and segmentation as well as data augmentation through skewing and Gaussian noise were performed to enhance the data quality. The feature extraction was performed using quite robust techniques like Hu and Zernike moments which have proven their efficacy in character recognition tasks.

The model was trained using three convolutional neural networks models Alexnet, Googlenet and Inception-V3 where Inception-V3 was the most accurate performing model. For the purpose of comparative performance evaluation with deep learning models, traditional machine learning models of SVM and KNN were also trained on a similar set of features.

Promising results were achieved using the proposed method for touch Brahmi character recognition and segmentation. Out of a total of 8014 images (963 touching characters) and obtained 100% accurate accuracy rate. The segment obtained after segmentation yielded 955 images of the identified

touching characters for 2187 individual characters identified with 99.16% accuracy. In comparison, SVM and KNN had lower identification rates, achieving 85% and 72% respectively.

Finally, the study ends with a discussion on the relevance of achieving high accuracy rates for identification and segmentation of touching characters and a suggestion of the next improvements that could follow if any errors were discovered [2].

Premi et al. [3] developed a more accurate model and a quicker computer-based method for identifying handwritten Brahmi letters, especially those written on palm leaves. To achieve their goal, the researchers have chosen to use an algorithm known as CNN (Convolutional Neural Network), which is known to be very successful at image related tasks. There are numerous steps in between forming a research methodology. In the first step, images of palm leaves are gathered, and some pre-processing techniques are used to enhance the quality of images. Then segmentation follows to separate the characters. Next, the CNN model was used to identify the characters.

It is found that this method reaching an accuracy of 73% can be done within about equal to a second. In short, the complete method involves the acquisition, pre-processing, segmentation, and application of CNN with an E-MNIST dataset. To accurately and effectively identify handwritten characters on palm leaf surface, a Selection Auto Encoder Decoder (SAED) technique was implemented for character-level recognition [3].

Preethi et al. [4] investigated the application of deep learning approaches to the automatic recognition of ancient Brahmi characters in handwritten, printed, and epigraphic Brahmi characters.

A dataset of 287 Brahmi characters was generated with around 210 samples for each class, to be used for training and testing. Note that to avoid bias, they collected data from different sources manually.

For character recognition, transferred learning strategies were applied using Alexnet, Googlenet and Inception-V3. At the end of the competition, a Hu-Zernike moments were used for the training of a traditional machine learning model. When it comes to the prediction of Brahmi characters, Inception V3 provides the highest accuracy, 97.5% while Alexnet and Googlenet show 93.6%, 95.4% accuracy respectively.

The recognition rates of conventional machine learning approaches were significantly lower than those obtained using deep learning methods, underscoring the strengths of these approaches for this character recognition task [4].

Aniket et al. described a technique for character recognition of Brahmi script. [5] in their study. The approach was based on Histogram of Orientated Gradients (HOG) features and a support vector machine (SVM) classifier with an error-correcting output codes (ECOC) model.

HOG descriptors: Images are preprocessed through steps like resizing and gradient computation to capture the changes in intensity in localized portions of the image (HOG features). It consists of 6000 characters from 32 classes of Brahmi with 80% data for training and 20% data for testing.

The overall recognition accuracy of the proposed system is 92.09%. The results ranged widely: Five classes scored 100% accuracy, 19 classes scored above 90% accuracy, and eight classes scored between 75 and 90% accuracy.

The results of the study show that the HOG feature extraction along with support vector machine classification achieves the best accuracy in the recognition of the Brahmi characters. However, a lower recognition rates for some characters was noticed that had the same shapes as others, which implies that if additional features were included, as in multi-feature recognition, it would increase the recognition accuracy [5].

Seba Susan et al. [6] introduced a novel system for the recognition of Devanagari script from digitized manuscripts.

The authors developed a two-phase classification scheme. Five convolutional neural networks (CNNs) were trained to extract features for each image quadrant (top-left, top-right, bottom-left, bottom-right and center) before constructing a feature vector. Then this vector was fed into a DNN having an output layer with 46 neurons representing 46 Devanagari character classes, and 3 hidden layers, each with 1000 neurons.

In order to evaluate the performance of the proposed method, the UCI Devanagari handwritten character dataset with 92,000 images belonging to 46 classes was used. This model achieved a stunning 98.39% classification accuracy, way ahead of all of the listed existing approaches of the study.

Some of the major challenges that are observed include a lack of rich datasets that can be used to train models, which is a common barrier for machine learning tasks. Due to this limited availability, the need for more extensive and diverse datasets has increased to improve efficiency in the identification

of handwritten scripts by AI models. Also, the interpretability of deep neural networks is still difficult, and it is hard to understand how the model makes decisions according to the content of images, even though the proposed method achieves a high level of accuracy [6].

Mahajan et al. [7] used a deep learning algorithm to recognize characters of ancient Indian MODI script. However, it still poses challenges for handwriting recognition because of its cursive similar glyph symbols which, if done manually, takes more time with more errors and confusion, since the MODI script was used for official administrations and literary writings in India between the 12th to 20th centuries.

Therefore, this research presents a CNN based system to automate the character recognition process which can help to increase the efficiency of transliteration and translation of ancient MODI script documents thus leads to the preservation of content.

The study framework is made up of the following steps:
1. Dataset Compilation: A dataset containing 5800 images of isolated characters from the MODI script was compiled from multiple sources.
2. Preprocessing: Images were preprocessed to correct deficiencies for recognizing the handwritten text.
3. Data Augmentation: One of the powerful techniques used is Data Augmentation, in which they performed some data augmentation techniques using MATLAB image Data Augmenter () function to boost the performance of the model and improve the dataset.
4. Model Training: The AlexNet model, which became famous for image recognition tasks, was used for training using MATLAB training Options () function

With this algorithm the MATLAB platform can achieve up to 89.72% accuracy after five epochs and 1015 iterations using AlexNet model [7].

Brindha et al. conducted an inquiry. [8] aimed at detecting and translating ancient Tamil Brahmi characters into modern Tamil scripts through advanced image processing and machine learning techniques.
This study involved a systematic recognition and transliteration approach to segment antique Tamil Brahmi's characters to new Tamil scripts. In the first stage, the ancient characters were photographed and preprocessed, including resizing, gray-scale conversion, noise reduction through Gaussian filtering, and binarization with the Otsu thresholding algorithm. Next, the images were cut into lines and characters. Feature extraction was carried out using a new combination of Zernike moments and zoning features, and a chi-square test was conducted to examine the independence of feature values. The retrieved features were used to train a neural network model using back propagation and gradient descent techniques for the purpose of character recognition.

The confusion matrix results showed that the proposed system achieved recognition accuracy and an error rate of 91.3% and 8.7% respectively. The images processed and their features extracted through the proposed methodology greatly helped in improving the character recognition performance of the model [8].

Devi et al. work [9] described a method to transform ancient Tamil Brahmi scripts into modern readable Tamil scripts utilizing image processing and deep learning.

The system proposal was made up of the following stages, carried out in this order:

1. Image Acquisition: After acquisition or scanning input images were stored in various formats (JPEG, TIFF, BMP, and PNG).
2. Image preprocessing: This step includes morphological processing, noise reduction, and transformation into grayscale image, which are performed for increasing the purity of image and accuracy of feature extraction.
3. Segmentation: words and characters from the processed images were obtained via Maximally Stable Extremal Regions (MSER) extraction.
4. Letter Recognition: The extracted letters were classified and recognized using deep learning techniques (specifically ResNet-50 classifier).
5. Letter Recognition: Discovered letters were visualized using MATLAB by contrasting them with similar Tamil letters recognized in modernity.

The system used community standards to reliably and quickly recognize and classify letters in different ways. Other techniques, such as stitching and blending, enhanced and led to the basic image processing of the individual images into a continuous surface. The methods were morphological and proceeded to noise filtering and grayscale conversion. Although so, when examining it more closely, deep learning has been essential in the categorization process. Ancient letters have been classified by the ResNet-50 model, which is famous for its recognition of complex neural networks. The categorization process was done by utilizing the distinguishing characteristics that were obtained

from the training images.

This technique involves associating the old Tamil Brahmi letters with usable Tamil scripts, after identifying them correctly.

The authors do not provide any specific performance metrics, but the approach of using ResNet-50 is indicative of better recognition and interpretation of historical texts as compared to traditional methods [9].
The authors mention no performance metrics but the concept behind ResNet-50 is an improvement towards analyzing and understanding historical texts compared to standard analysis techniques [9].

This study conducted by Saxena et al. [10] to improve the accuracy of the preservation of historic handwritten Devanagari documents. They achieved the phenomenal accuracy of 97.4% by integrating both CNN method along with the Hierarchical Optical Character Recognition (HOCR) methods.

Using the standardized HOCR format, which represents results of optical character recognition in a hierarchical structure, it's possible to extract text information and layout from digital documents. After extracting text, the CNN model would be trained for Devanagari character recognition. Currently, a certain category of deep learning methods is referred to as the convolutional neural networks (CNNs), and it is well known for its high-performance image recognition capabilities. In this paper, neural network technology is used to train the model for handwritten Devanagari letters recognition using the dataset. The aim of the training was to identify the unique patterns & reports which make each character different from the other. Due to this learning mechanism, the model was able to recognize and classify the characters that were written casually.

The study emphasizes the importance of preserving our historical records for future generations, as these documents contain rich details of our past. Going digital with historical records ensures they'll be permanent; more records won't be lost as time passes and saves time for archivists. Hierarchically, intra-document consistency through CNN neural network methods as another layer of improvement highlights the potential feasible recognition rates and individual caveats of the techniques for enhancing text retrieval in ancient manuscripts. The use of a large dataset [10] also contributes to the results being accurate.
The research article by Singh et al [11] is mainly concerned with the idea of dividing a document image into its text entities (e.g. words, text lines, characters, sub-characters). This technique is specifically for Handwritten Brahmi Script.

The proposed approach consists of many main stages. The text was first segmented into line by line with the MATLAB Line Segmentation function. The process of segmenting a text from document images is performed step by step. First, Text lines are divided into Words, and then Words are broken down into individual Characters. At times when it became evident that certain characters were similar to others, some rule was applied to separate them out from another character. Regrettably, the rate at which the method for segmentation accuracy has been given is not stated in it. The main focus of this work was to identify and disentangle different units constituting a document image. This paper mainly deals with segmentation techniques especially emphasizing an extensive study about Brahmi script recognition [11].

An investigation performed by Nilupuli Wijerathna et al. [12] related to the determination of Brahmi ancient characters discovered in inscriptions dated ranging from the 3rd century B.C to the 1st century A.D. A multi-phase process was then proposed to accurately detect and interpret these inscriptions.

First, the process consists of the noise removal and the enhancement of the inscription picture using image processing methods to obtain element segmentation and binary version. Once the segmentation work was done the next challenge was tackling broken or incomplete characters. For this study, a model was developed and trained with the help of deep learning techniques, mainly CNN, for attaining high accurate quality classification of Brahmi script character. The approach focused on reading and explaining the significance of the inscriptions rather than identifying what letters were there and dating them. These periods were annotated and stored in a database for later use.

The system was basically designed to read ancient Brahmi letters using an image-processing and deep-learning-based approach followed by the subsequent natural language processing for interpretation and translation. Although both accuracy of the CNN model and the overall system are unknown because of the complexity of everything here, the comprehensive approach was able to infer the level of accuracy in terms of how close that the predictions towards emancipation of an ancient text concept.

In summary, the study report outlined an approach for identification and understanding of ancient Brahmi script that was found on inscriptions. Alongside tasks related to identifying and transcribing the ancient engravings as Sinhala, the work was further based on pre-processing images, segmenting characters and detecting Brahmi script using CNNs. Using the capabilities of image processing, deep learning and natural language

processing methods, it performed recognition, interpretation and translation tasks effectively [12].

Wickramarathna et al. [13] addresses character-level processing and conversion of Brahmi characters into Sinhala words in the post-processing stage after OCR recognizes the characters. The method designed in this paper focuses on enhancing the recognition of words, especially for Brahmi and Sinhala, and therefore handling OCR erroneous translations.

OCR corrections through Neural Networks based NLP. The OCR error-correcting module consists of multiple elements:

1. Morphological: This is where the array of Brahmi characters along with the potential errors made during writing or the changes implemented.
2. Permutations: The generation module produces alternative combinations of these problems to address different types of repair methods.
3. Minimal Edit Distance Approach: The minimal edit distance computation will tell you, by comparing the error and correction, about more likely repairs for each problem.
4. Best Proposal Model: the best proposal model takes the least edit distance and other parameters into consideration to find the best correction proposal.
5. Develop Word Patterns: The approach applied bigram, trigram models and generated an approximate fluid order of words in the resulting Sinhala word using the adjusted order of Brahmi letters. Its technology boasts an accuracy rate of between 86% and 91%. The system's approach includes adapting Brahmi script characters to the correct Sinhala words utilizing linguistic models, Optical Character Recognition (OCR) correction process, and NLP techniques.

This holistic technique successfully addresses OCR mistakes and enhances word location recognition precision [13].

Suganya et al. [14] discussed the recognition of the Ancient Tamil alphabet based on exploratory study that is uses for point identify shapes and Hough drift, Group Search Optimization, selected point & Firefly algorithm. For the recognition phase, algorithms Neural Network, J48, naive baes and KNN were used and according to researches, the accuracy ranges from 91% to 94%. As a first step, one must read the ideal script of Indian Tamil derivation in order to come up with an image of the script. Apply median filtering to improve the image quality and reduce noise. The photograph is pre-processed and converted to binary.

Characters are subsequently excluded from the script and granted additional time for processing. Using the shape and Hough transform algorithms to extract the features, relevant details on the curvature and shape of the characters are then extracted. We then use Firefly and Graph Search Optimization algorithms for selecting the most discriminative and relevant qualities for classification task.

And finally, the classification is done via Neural Network and J48 (C4. decision tree, Naïve Bayes and KNN All these algorithms use retrieved attributes for grouping the characters.

This article describes a method that achieves the above-complex task to read ancient Tamil script by using a few techniques such as preprocessing segmentation, position detection, selection, and classification. They have also successfully and very accurately reported recognition results using of shape and Hough transform techniques combined with advanced point selection algorithms [14].

In their study work, Gautam et al. [15] presented a Handwriting (printed and handwritten) for Brahmi characters recognition technique. The recognition of Brahmi script involves four elementary processes namely, distinguishing the characters or character classes, extracting useful features, segmenting the characters and preprocessing the image.

First, Pre-processing operations are performed on the Brahmi characters in an image. These are for reducing the noise and artifacts that are produced when doing the recognition.

To separate each character for future analysis and feature extraction, the image was divided. This approach utilizes geometric components to analyse the attributes of Brahmi characters. After gaining the attributes, a range of tagged Brahmi character examples were used to sort the characters and divide them into relevant categories using an OCR method.

It achieved accuracy rates of 94.90% for printed consonants, 93.30% for printed vowels, 89.55% for handwritten consonants, and 91.69% for handwritten vowels when tested with printed and handwritten Brahmi letters [15].

III. Summary of Existing Work

To address the complexities of the ancient script, academicians have investigated several methods in the field of OCR for the Brahmi script. CNNs are a popular method that frequently defines Brahmi characters correctly. The VGG16 architecture, which achieved over 96% accuracy in script categorization, highlights the potential of deep learning in this field. Brahmi script pre-processing and picture segmentation have been extensively studied. To improve the dataset quality and increase the

recognition accuracy, they have applied skewing, noise reduction, grayscale conversion, and Gaussian noise. Two feature extraction techniques-Hu and Zernike moments-have been successfully used to extract the distinctive characteristics of Brahmi characters.

Additionally, in terms of the area of deep learning and ensemble classification methods models such as Alexnet, Googlenet, and Inception-V3 have been trained on large datasets to recognize Brahmi characters and Inception-V3 is an accurate performer. In order to compare Traditional machine learning techniques, like Support Vector Machines (SVMs) and K-Nearest Neighbours (KNN), have also been utilized, and the results demonstrate that deep learning techniques are better when applied to this particular problem.

In addition, the post-processing of optical character recognition (OCR) has focused on eliminating errors and refining the recognition of Brahmi writing. Brahmi characters have been translated into modern language meaningful text with the use of a variety of techniques (e.g., Morphological analysis, permutation generation, minimum edit distance and bigram, trigram models).

In general, the previous OCR studies for the Brahmi characters have made tremendous strides, emphasizing using complex image processing and AI methods to increase performance.

TABLE 1: SUMMARIZING EXISTING WORK ON CHARACTER RECOGNITION USINGDIFFERENT MODELS

Author	Script Used	Method/Technique Used	Result	Research Gaps
Vincen et al. [1]	Brahmi	VGG16 architecture with CNN, preprocessing (zero padding, resizing, grayscale, binarization), data augmentation	96% accuracy, 98% recall and precision	Need for more extensive and diverse datasets
Aniket et al. [2]	Brahmi	Ensemble classification technique, preprocessing (noise removal, grayscale, segmentation), feature extraction (Hu and Zernike moments), CNN models (Alexnet, Googlenet, Inception-V3)	100% accuracy in identifying touching characters, 99.16% accuracy in segmentation	Lower accuracy for non-touching characters
Premi et al. [3]	Brahmi	CNN for character identification on palm leaf surfaces, preprocessing (data acquisition, preprocessing, segmentation)	Approximately 73% accuracy	Need for higher accuracy and faster computational methods
Preethi et al. [4]	Brahmi	Transfer learning utilizing AlexNet, GoogleNet, and Inception-V3, conventional machine learning methods trained on Hu-Zernike moments.	Inception-V3: 97.5% accuracy, Alexnet: 93.6%, Googlenet: 95.4%	Lower accuracy for traditional machine learning models
Aniket et al. [5]	Brahmi	Histogram of Oriented Gradientsfeatures, Support Vector Machineclassifier enhanced by Error-Correcting Output Codes (ECOC) model	92.09% overall recognition accuracy	Lower accuracy for characters with similar shapes
Seba Susan et al. [6]	Devanagari	Two-phase classification system using CNNs and a deep neural network	98.39% accuracy	Interpretability of deep neural networks
Mahajan et al.[7]	MODI	CNN for character recognition. The AlexNet model was utilized for training the system.	89.72% accuracy	Need for more extensive and diverse datasets
Brindha et al. [8]	Tamil Brahmi	Preprocessing (resizing, grayscale conversion, Gaussian filtering, binarization), feature extraction (Zernike moments and zoning features), neural network model trained via backpropagation and gradient descent	91.3% recognition accuracy	Higher error rate (8.7%)
Devi et al. [9]	Tamil Brahmi	Image processing (grayscale conversion, noise removal, morphological processing), segmentation (MSER method), deep learning (ResNet-50 classifier)	Not specified	Need for specific accuracy metrics
Saxena et al. [10]	Devanagari	Combination of HOCR and CNN techniques	97.4% accuracy	Need for more robust datasets
Singh et al. [11]	Brahmi	Line segmentation using MATLAB, further segmentation into words and characters	Not specified	Need for accuracy metrics
NilupuliWijerathna et al. [12]	Brahmi	Image processing, deep learning (CNN), NLP for translation into modern Sinhala	Not specified	Need for specific accuracy metrics and interpretability
Wickramarathna et al. [13]	Brahmi	NLP with OCR and OCR correction techniques, morphological analysis, permutations generation, minimum edit distance approach, best suggestion module, generate word sequences	86% to 91% accuracy	Need for higher accuracy and faster computational methods
Suganya et al. [14]	Tamil	The methodologies employed for feature	91% to 94%	Lower accuracy for some

		extraction include the Shape and Hough transform techniques. Group Search Optimization and Firefly algorithm are used for feature selection. Furthermore, Neural Networks, J48, Naïve Bayes and K-Nearest Neighbours (KNN) are used as the classifiers.	accuracy	algorithms
Gautam et al. [15]	Brahmi	Image preprocessing, character segmentation, feature extraction (geometric features), character classification using OCR	91.69% for handwritten vowels, 89.55% for handwritten consonants, 93.30% for printed vowels, 94.90% for printed consonants	Need for higher accuracy for handwritten characters

The comparative analysis for several methodologies applied for Brahmi text extraction and recognition has been presented in Table I. These methodologies include various aspects, starting from image processing and preprocessing and including feature extraction as well as classification and deep learning methods.

IV. Conclusion

Based on various Optical Character Recognition (OCR) systems developed so far for High Quality Brahmi script extraction and recognition, it is apparent that AI and ML play a significant role in unearthing the historical and cultural treasures found in ancient documents. These complexities and degeneracies have prompted the development of new techniques to rectify this problem in the Brahmi writing system. Innovative developments in image preprocessing, segmentation and character recognition have emerged from these techniques. Deep learning-based models, especially CNNs, have been proven to significantly boost both the efficiency and accuracy of the OCR systems whereas post-processing methods have helped further improve modern language interpretation of Brahmi script. It is possible that with advancements in technology, the character recognition systems may reach more accuracy and sophistication. This will mean that our cultural heritage will be preserved and available for future generations.

REFERENCES

[1] Vincen, Vincen, and Samsuryadi Samsuryadi. "Brahmi Script Classification using VGG16 Architecture Convolutional Neural Network." Computer Engineering and Applications Journal 11.2 (2022): 127-136.

[2] A.S., N. & Mali, S. (2021). Identification and Segmentation of Touching Brahmi Characters From Degraded Digital Estampage Images Using Ensemble Classifier. Indian Journal of Computer Science and Engineering, 12(6):1722–1733. DOI: 10.21817/indjcse/2021/v12i6/211206110.

[3] P.Premi a ,R.Madhumithab, N.R.Raajan, "CNN based Digital alphanumeric archaeolinguistics apprehension for ancient script detection", Turkish Journal of Computer and Mathematics Education, Vol.12 No.6(2021), 5320-5326.

[4] P. P and M. H. R, "Intelligent Recognition of Ancient Brahmi Characters using Transfer Learning," 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2023, pp. 1-6, DOI: 10.1109/ICAECT57570.2023.10118090.

[5] A. S. Nagane, C. H. Patil and S. M. Mali, "Classification of Brahmi script characters using HOG features and multiclass error-correcting output codes (ECOC) model containing SVM binary learners," 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bengaluru, India, 2023, pp. 448-451, DOI: 10.1109/IITCEE57236.2023.10091084.

[6] Susan, S., & Malhotra, J. (2020). Recognising Devanagari Script by Deep Structure Learning of Image Quadrants. DESIDOC Journal of Library & Information Technology, 40(05), 268-271. https://doi.org/10.14429/djlit.40.05.16336.

[7] Mahajan, Kirti & Tajne, Niket. (2021). An Ancient Indian Handwritten Script Character Recognition by Using Deep Learning Algorithm.

[8] Brindha, S. & Bhuvaneswari, S.. (2021). Repossession and Recognition System: Transliteration of Antique Tamil Brahmi Typescript. Current Science. 120. 654. 10.18520/cs/v120/i4/654-665.

[9] P. D. Devi and V. Sathiyapriya, "Brahmi Script Recognition System using Deep Learning Techniques," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 1346-1349, DOI: 10.1109/ICIRCA51532.2021.9544978.

[10] N. Saxena and S. Chauhan, "Transformation of handwritten Devnagari script into word editable form using CNN," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2020,734-738, DOI:10.1109/ICACCCN51052.2020.9362824.

[11] Singh, Ajay P., and Ashwin Kumar Kushwaha. "Analysis of Segmentation Methods for Brahmi Script." DESIDOC Journal of Library & Information Technology 39, no. 2 (2019), 109-116. https://doi.org/10.14429/djlit.39.2.13615.

[12] K. A. S. A. Nilupuli Wijerathna et al., "Recognition and translation of Ancient Brahmi Letters using deep learning and NLP," 2019 International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 2019, pp. 226-231, DOI: 10.1109/ICAC49085.2019.9103340.

[13] S. Wickramarathna and L. Ranathunga, "Data Driven Approach to Brahmi OCR Error Correction and Sinhala Meaning Generation from Brahmi Character Array," 2019 19th International Conference on Advances in ICT for

Emerging Regions (ICTer), Colombo, Sri Lanka, 2019, pp. 1-6, DOI: 10.1109/ICTer48817.2019.9023763.

[14] T. S. Suganya and S. Murugavalli, "Feature selection for an automated ancient Tamil script classification system using machine learning techniques," 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), Chennai, India, 2017, pp. 1-6, DOI: 10.1109/ICAMMAET.2017.8186731.

[15] Gautam, Neha & Chai, Soo See. (2017). Optical Character Recognition for Brahmi Script Using Geometric Method.