RECOGNITION OF TAMIL ANCIENT CHARACTERS AND INFORMATION RETRIEVAL FROM TEMPLE EPIGRAPHY USING MACHINE LEARNING

Mathumitha A, Sujith A, Visali A, Sineka V

[1]*Mathumitha A, Assistant Professor, Bannari Amman Institute of Technology, Tamil Nadu, India*
[2] *Sujith A, Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India*
[3] *Visali A, Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India*
[4] *Sineka V, Computer Science and Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*

ABSTRACT

A modern method of identifying the characters of ancient Tamil inscription. This paper presents an advanced approach to deciphering the characters found in ancient Tamil inscriptions, with a specific focus on the Brahmi type of Tamil characters. Unlike modern Tamil, which is widely understood, these characters, dating back to the 14th century, pose a significant challenge in interpretation. These inscriptions hold invaluable information concerning the lives of kings, their societal contributions, as well as insights into the field of medicine during that era. Consequently, this research paper introduces a novel method for extracting meaningful information from these inscriptions and written scripts by employing a Hybrid Machine Learning Model that combines Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Through this innovative approach, the ancient characters can be accurately recognized and subsequently translated, leading to a deeper understanding of historical and medical knowledge encoded within these valuable sources.

Keywords:- *Brahmi characters - Hybrid machine learning model - Convolutional Neural Network - Recurrent Neural Network.*

I. INTRODUCTION

Ancient Tamil inscriptions are invaluable historical artifacts that offer a unique window into the lives of kings and the advancements made in medicine during the 14th century. These inscriptions, written in the Brahmi type of Tamil characters, present a significant challenge for interpretation due to their distinct form, which differs from modern Tamil. Dating back to the 14th century, these ancient Tamil inscriptions provide essential insights into the cultural, social, and political landscape of the time. They offer glimpses into the lives of the ruling kings, their contributions to society, and the state of medical knowledge and practices prevalent during that era. These inscriptions represent an extensive collection of primary sources, shedding light on historical events, societal structures, and advancements in various domains, particularly in the field of medicine. The Brahmi type of Tamil characters used in these inscriptions present a formidable challenge for interpretation. These characters significantly differ from the modern Tamil script, both in their visual appearance and contextual usage. As a result, deciphering and comprehending these ancient characters requires specialized knowledge and expertise. The distinctiveness of the Brahmi type of Tamil characters poses difficulties for individuals familiar only with the modern Tamil script to grasp their meaning and significance. Decoding these characters demands a multidisciplinary approach that encompasses linguistic analysis, historical context, and the application of advanced machine-learning techniques. The Brahmi type of Tamil characters, inscribed on various surfaces such as stone, copper plates, and pottery, may have been subject to erosion and damage over time. This degradation further complicates the process of interpretation making it necessary to employ innovative methods and technologies to enhance legibility and extract meaningful information. Previous studies on character recognition and machine learning techniques have primarily focused on modern scripts and languages. While these approaches have demonstrated success in contemporary contexts, applying them to ancient. Tamil characters pose unique challenges. The distinctive visual characteristics, the lack of standardized forms, and the evolution of the Tamil script over time necessitate a tailored approach specifically designed for the Brahmi type of Tamil characters.

To address these challenges, this research proposes a Hybrid CNN-RNN Model for character identification and interpretation. This model combines the strengths of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to effectively recognize and understand the ancient Tamil characters. The CNN component is responsible for extracting visual features from the inscriptions, such as lines, curves, and shapes. By learning hierarchical representations of these visual patterns, the CNN enables the model to capture the unique characteristics of the ancient characters. In parallel, the RNN component of the model focuses on understanding the sequential nature of the character sequences. It captures the contextual dependencies present within the inscriptions, allowing for a deeper comprehension of the linguistic and textual nuances. By combining these two components, the Hybrid CNN-RNN Model offers a holistic and comprehensive approach to character identification, enhancing the accuracy and interpretability of the results.

The success of the proposed model relies on the availability of a diverse and representative dataset of ancient Tamil inscriptions. Collecting such a dataset is a crucial step in ensuring the model's effectiveness and generalizability. The dataset should include inscriptions from various sources, time periods, and geographical locations, allowing for a comprehensive exploration of the different styles, variations, and challenges presented by the Brahmi type of Tamil characters. Moreover, preprocessing techniques play a vital role in optimizing the dataset for training the Hybrid CNN-RNN Model. These techniques encompass image enhancement to improve legibility, noise removal to mitigate the impact of erosion and degradation, and normalization procedures to ensure consistency and comparability across the dataset. By employing appropriate preprocessing techniques, the model can better handle the complexities inherent in the ancient inscriptions, facilitating more accurate character recognition and interpretation. Image recognition, in the context of machine vision, is the ability of software to identify objects, places, people, writing, and actions in images. It is achieved in combination with a camera and artificial intelligence technique. Software for image recognition requires deep machine learning. Image recognition algorithms can function on 3D models.

There are many types of image formats that are available which on conversion may lead to loss of data. The main aim of image conversion is to retrieve the information in the image without any loss.

II. RELATED WORK

Various efforts have been taken in translating ancient Tamil which have been successfully proposed for Brahmi type of letters. The other two important format of ancient Tamil are Vattazhuthu and Grantha which has been posing great difficulty for translation due to lack of a database for these letters. Grantha letters have also been translated with minimum training in the database. Although all the efforts did not yield full efficiency for translating the letters phases of previous works done are continued to improve the efficiency. The proposed method is an effort to translate Brahmi letters.

III. PROPOSED METHOD

In this method, the input image is first gray-converted and then pre-processed using the median filter to remove the noises. Then the contour-let transform is used to enhance the image with respect to the center object which is followed by the morphological operation. It performs four operations namely dilation, erosion, filling and taking details of regional properties. Using the information provided by the regional properties textural feature extraction is done for each segmented image. The feature of the dataset is also extracted and trained using the Back Propagation Neural Network method and the features are compared and the Fig.1 illustrates the methodology.

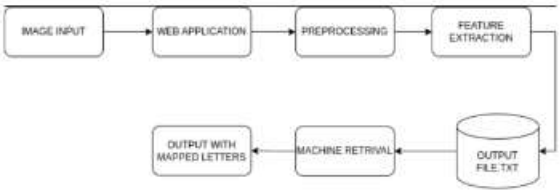

Fig 1:- Block diagram of the proposed method

A. Image Enhancement:

Noise reduction: Apply filters and algorithms to remove noise and artifacts from the images of the ancient Tamil inscriptions. This helps improve the legibility of the characters. Contrast adjustment: Enhance the contrast of the images to bring out the details and improve the visibility of the characters against the background. Sharpening: Use techniques like unsharp masking or high-pass filtering to enhance the sharpness and clarity of the characters.

B. Segmentation:

Character segmentation: Separate individual characters within the inscriptions by detecting spaces or boundaries between them. This step is crucial for isolating characters and preparing them for further analysis. Line segmentation: Identify and extract individual lines of text within the inscriptions to facilitate a more structured analysis and recognition process

C. Normalization:

Size normalization: Resize the characters to a standard size to ensure consistency across the dataset. This step helps in removing variations caused by differences in the original size of the characters in the inscriptions. Orientation correction: Rectify the orientation of the characters to ensure they are properly aligned. This can involve rotation techniques to eliminate any skew or tilt present in the images.

D. Background Removal:

Gravy level co-occurrence matrix (GLCM) is used to ex-homogeneity. The Matrix is formed using the function gray comatrix and the statistics are calculated using the function graycoprons.

E. Neural Network Classifier:

The type of neural network used is Probabilistic Neural Network (PNN). It calculates the probability distribution function of each class and compares it with the input class and then allocates the highest probability rate for the output.

Fig 1: - Output of single letter

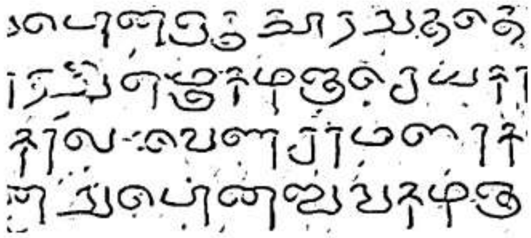

Fig 2:- Output image of group of letters

IV. CONCLUSION

This work concentrates on translating the Vattezhuthu style of ancient Tamil. Further, the work can be improved by adding more details to the database using handwritten scripts and improving the algorithm for image recognition for efficient translation of the whole image. The work can be further developed by combining the works done for all the three format of ancient Tamil letters to create an application which would be useful for people to understand the inscriptions of Tamil language.

REFERENCES

[1]. Yoshinari Tsuzuki, Kei Sawada, Kei Hashimoto, Yoshihiko Nankaku and Keiichi Tokuda- "Image Recognition based on Discriminative Models using features generated from separable lattice HMMs"(ICASSP-2017)

[2]. R.Giridharan, E.K. Vellingiriraj and P. Balasubramanie- "Identification of Tamil Ancient Characters and Information retrieval from temple epigraphy using image zoning" (ICRTIT 2016).

[3]. M.Mahalakshmi, Malathi Sharavanan- "Ancient Tamil Script and Recognition and Translation using LabVIEW" (ICCSP 2013).

[4]. K. Sawada, A. Tamamori, K. Hashimoto, Y. Nankaku, and K. Tokuda, "Bayesian approach to image recognition based on separable lattice hidden Markov models," IEICE Transactions on Information and Systems, vol.E99-D, no.12, pp3119–3131, 2016.

[5]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma,Z. Huang, A. Karpathy, A. Khosla, M.Bernstein, A.C. Berg, and L. Fei-Fei,"ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision,vol. 115, no. 3, pp. 211–252, 2015.

[6]. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R.B. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Con-volutional architecture for fast feature embedding," ACM international conference on Multimedia, pp. 675–678, 2014. [7]. S. Rajkumar and Dr. Subbiah Bharathi - "Century Identification and Recognition of Ancient Tamil Character Recognition", IJCA(0975-8887), Volume 26- No.4, July 2011.