PREHISTORIC TAMIL SCRIPTURE RECOGNITION FROM STONE IMAGES

**A MINI PROJECT REPORT**

***Submitted by***

**DHARSHINI PRIYA G (1920110008) LIZZY ANGELIN ALPHONSA G (1920110023)**

**MOHSIN TAJ M (1920110029)**

***in partial fulfillment of the award of the degree of***

# BACHELOR OF TECHNOLOGY

**IN**

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

**SONA COLLEGE OF TECHNOLOGY, SALEM-5**

**(Autonomous)**

# ANNA UNIVERSITY: CHENNAI 600 025

NOVEMBER 2023

# ANNA UNIVERSITY: CHENNAI 600 025

# BONAFIDE CERTIFICATE

Certified that this project report “**PREHISTORIC TAMIL SCRIPTURE RECOGNITION FROM STONE IMAGES”** is the bonafide work of **“DHARSHINI PRIYA G (1920110008) LIZZY ANGELIN ALPHONSA G (1920110023) MOHSIN TAJ M (1920110029)”**who

carried out the project work under my supervision

**SIGNATURE SIGNATURE**

Dr . J. Akilandeswari Mr. David Samuel Azariya S

**HEAD OF THE DEPARTMENT SUPERVISOR**

Professor Assistant Professor

Department Of Information Technology Department Of Information Technology Sona College of Technology Sona College of Technology

Salem- 636 005. Salem- 636 005.

Submitted for Project viva voce examination held from 16.11.2023 to 18.11.2023

|  |  |
| --- | --- |
| **INTERNAL EXAMINER** | **EXTERNAL EXAMINER**  **ii** |

# ACKNOWLEDGEMENT

First and foremost, we thank to **power of the almighty** for showing us inner peace and for all blessings. Special gratitude to our parents, for showing their support and love always.

We express our sincere thanks to Chairman **Sri. C. VALLIAPPA** and principal **Dr.S.R.R.SENTHILKUMAR** for providing adequate facilities to complete the project.

We are immensely grateful to our Head of the Information Technology Department, **Dr. J. AKILANDESWARI for** the continued encouragement to complete the project.

We express our heartfelt thanks to our project supervisor **Mr. David Samuel Azariya S** for his valuable guidance and fruitful discussions throughout the project work.

We feel proud to share this success with all our department Faculty, Staff members, and friends who helped directly and indirectly to complete this project successfully.

**iii**

ABSTRACT

Tamil is a script-based language with characters that are written in various styles and fonts. Recognizing Tamil letters from images is a challenging task due to the high similarity and variability among the characters. The intricate challenge of Prehistoric Tamil character recognition from stone images, which is inherently more complex due to the characters' intricate shapes, numerous holes, loops, and curves. Despite previous efforts, significant challenges persist in this domain. This study proposes an innovative approach employing Deep Neural Networks (DNN) to tackle the issue of recognizing prehistoric Tamil characters. The image preprocessing pipeline includes binarization, denoising, character segmentation, and size normalization, which are crucial steps for enhancing recognition accuracy.Moreover, feature extraction plays a pivotal role, and in this context, both Fourier feature mappings enriching the recognition process with valuable information. Character classification is performed utilizing Back Propagation DNN, while optimization of neural networks leverages the Simplex method during backpropagation. The proposed system strives to provide an improved approach to address the inherent complexities of prehistoric Tamil character recognition, offering a promising avenue for historical and archaeological research applications

**iv**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** | **iv** |
|  | **LIST OF FIGURES** | **vi** |
| **1.** | **INTRODUCTION** |  |
| Objective of the Project 7  Literature Survey 8  Existing Systems 9  Hardware & Software Requirements 11 | | |
| **2.** | **DESIGN & IMPLEMENTATION** |  |
| Proposed System 15  Architecture Design 20  Modules 23  Modules Description 26  Screenshots & Results 34 | | |
| **3.** | **CONCLUSION** | 37 |
|  | **REFERENCE** | 49 |

**v**

# LIST OF FIGURES

|  |  |
| --- | --- |
| **CHAPTER 2** | **PageNo.** |
| Fig 2.1 – Architecture Design Fig 2.2 – Input Image | 20  35 |
| Fig 2.3 – Denoised Image Fig 2.4 – Boxed Image  Fig 2.5 – Segmented Character Image  Fig 2.6 – Model Accuracy | 35  36  36  37 |

**LIST OF ABBREVIATIONS**

DNN - Deep Neural Networks

OCR - Optical Character Recognition CNN - Convolutional Neural Networks

# vi

# CHAPTER 1 INTRODUCTION

Recognizing Tamil characters, especially prehistoric ones from stone images, is a complex task due to their intricate shapes. This study introduces a novel approach using Deep Neural Networks (DNN) for improved recognition. The image preprocessing pipeline includes crucial steps like binarization, denoising, character segmentation, and size normalization to enhance accuracy. Feature extraction is pivotal, with Fourier feature mappings and Zernike moments providing valuable descriptors. Character classification employs Back Propagation DNN, optimized using the Simplex method. This system aims to address the challenges of prehistoric Tamil character recognition, offering promise for historical and archaeological research.

## OBJECTIVES OF THE PROJECT

The primary objective of this study is to develop a robust recognition system for prehistoric Tamil characters from stone images, a challenging task due to their intricate shapes and high variability. Through the application of Deep Neural Networks (DNN), we aim to enhance recognition accuracy significantly. Our approach encompasses a comprehensive image preprocessing pipeline that includes binarization, denoising, character segmentation, and size normalization. Additionally, we employ advanced feature extraction techniques, such as Fourier feature mappings and Zernike moments, to enrich the recognition process with valuable information. Character classification is carried out using Back Propagation DNN, with optimization leveraging the Simplex method during backpropagation. The ultimate goal is to provide an effective solution for the complexities associated with prehistoric Tamil character recognition, offering promising avenues for historical and archaeological research applications.

## LITERATURE SURVEY

**Recognition of Tamil Characters from Stone Inscriptions by N. M. Balasubramanian and S. Ramakrishnan**

Description: This paper is a foundational work in the field of recognizing Tamil characters from stone inscriptions. It addresses the challenges in deciphering characters from historical stone inscriptions and discusses techniques for image preprocessing, character segmentation, and recognition. The authors emphasize the cultural and historical significance of preserving and interpreting these inscriptions. The paper lays the groundwork for subsequent research in this area. **Character Recognition from Tamil Stone Inscriptions by S. K. Ranjani and**

## M. Pushpa.

Description: This paper focuses on the technical aspects of character recognition from Tamil stone inscriptions. It delves into image processing techniques, particularly binarization and denoising, to enhance image quality. The authors explore methods for character segmentation, separating individual characters from complex backgrounds. Their work is essential for understanding the key image processing challenges in this domain.

## Efficient Tamil Character Recognition from Stone Inscriptions R. Shanthi and A. Malathi

Description: This paper introduces an efficient approach to Tamil character recognition from stone inscriptions, where efficiency is critical for processing large volumes of historical inscriptions. The authors highlight the application of deep learning models, particularly convolutional neural networks (CNNs), to achieve high recognition accuracy. Their work showcases the potential of modern machine learning techniques for this task.

## Automatic Recognition of Tamil Characters from Stone Engravings T. Velmurugan and K. Thanushkodi

Description: This paper is dedicated to the development of an automatic recognition system for Tamil characters from stone engravings. It employs a combination of image processing techniques and machine learning algorithms for character segmentation and recognition. The paper provides practical insights into building systems for character recognition in the context of historical engravings.

## Preserving Cultural Heritage: A Survey on Tamil Character Recognition from Stone Images by G. Suresh and V. Priyadarsini

Description: This survey paper offers a comprehensive overview of the research landscape in Tamil character recognition from stone images. It emphasizes the importance of preserving cultural heritage through accurate recognition of inscriptions. The paper covers a wide range of approaches, both traditional and modern, highlighting the challenges and advancements in this field.

## Deep Learning Approaches for Tamil Character Recognition from Stone Inscriptions by R. Karthik and S. Meena

Description: This paper explores the application of deep learning approaches, particularly CNNs, for recognizing Tamil characters from stone inscriptions. It delves into the architecture design, training, and evaluation of deep learning models for this specific task. The authors discuss the advantages of deep learning in achieving high recognition accuracy.

## EXISTING SYSTEMS

Tamil character recognition from stone images, especially those based on advanced technologies like Convolutional Neural Networks (CNNs). Character recognition from stone inscriptions, especially in the context of prehistoric scripts, often relies on traditional methods involving manual transcription and expert interpretation, as mentioned earlier. However, there might be related

systems and technologies in the field of character recognition, historical script analysis, or cultural preservation. Here are some relevant existing systems and technologies:

**Optical Character Recognition (OCR) Software:** OCR software, such as Tesseract, is commonly used for character recognition from printed text and documents. While not specific to stone inscriptions, OCR technology can be adapted for recognizing characters from images. Tesseract OCR is an open-source Optical Character Recognition (OCR) engine developed by Google. It is one of the most popular and widely used OCR engines globally. Tesseract is primarily designed for recognizing printed and handwritten text in various languages and scripts. While its main focus is on modern printed text, it can be adapted and customized for recognizing prehistoric Tamil characters from stone images or other script-based characters.

**Historical Script Recognition Systems:** There are existing systems and research in the field of historical script recognition, which may include the recognition of ancient scripts like Tamil. These systems often involve image processing and machine learning techniques.

**Archaeological Databases:** Some archaeological organizations and institutions maintain databases of inscriptions and historical artifacts, including stone inscriptions. These databases may include scanned images and transcriptions.

**Cultural Heritage Preservation Tools:** Various tools and technologies are employed for the digital preservation of cultural heritage, including historical inscriptions. This might involve 3D scanning, high-resolution imaging, and archiving systems.

**Library and Archival Systems:** Libraries and archives often use digital cataloging systems for historical documents and inscriptions. While not focused on stone inscriptions, these systems can serve as a reference for character recognition.

## HARDWARE AND SOFTWARE REQUIREMENTS

Recognizing prehistoric Tamil characters from stone images using advanced technologies like Convolutional Neural Networks (CNNs) involves a combination of hardware and software components. Below are the key components for such a system:

## Hardware Components:

**Computing Hardware:** High-performance computing systems are essential for running deep learning models efficiently. This may include multi-core processors, Graphics Processing Units (GPUs), or specialized AI accelerators.

**High-Resolution Scanners or Cameras:** To capture clear and detailed images of stone inscriptions, high-resolution scanners or cameras are needed. These devices should be capable of capturing images with high fidelity.

**Display and Visualization Tools:** High-quality displays are needed for viewing and interpreting the recognized characters. Visualization tools aid researchers in understanding and working with the data. High-quality displays are needed for viewing and interpreting the recognized characters. Visualization tools aid researchers in understanding and working with the data. Display and visualization tools are essential for interpreting and presenting data and images in a comprehensible and visually appealing manner. These tools allow users to explore, analyze, and communicate information effectively. In the context of image processing and computer vision, display and visualization tools are crucial for visualizing results, sharing findings, and supporting research efforts.

## Software Components:

**Image Processing Software:** Image processing tools are used to enhance the quality of stone inscription images. This includes techniques like noise reduction, contrast adjustment, and image preprocessing.

**Convolutional Neural Network (CNN) Frameworks:** Deep learning frameworks like TensorFlow, PyTorch, or Keras are essential for developing and training CNN models for character recognition. These frameworks provide pre- built neural network layers and training capabilities. Convolutional Neural Networks (CNNs) are a category of deep learning models designed to process and analyze visual data, making them particularly well-suited for tasks in computer vision and image processing. These networks have revolutionized the field, enabling significant advancements in object detection, image classification, facial recognition, medical image analysis, and more. CNN frameworks are essential tools for building and deploying these models, simplifying the development process and improving their efficiency.

**OCR Software**: Optical Character Recognition (OCR) software can be used for initial character extraction from images. While not specific to stone inscriptions, OCR tools can be a starting point for character recognition. Tesseract OCR is an open-source Optical Character Recognition (OCR) engine developed by Google. It is one of the most popular and widely used OCR engines globally. Tesseract is primarily designed for recognizing printed and handwritten text in various languages and scripts. While its main focus is on modern printed text, it can be adapted and customized for recognizing prehistoric Tamil characters from stone images or other script-based characters.

**Data Annotation Tools**: Tools for annotating and labeling recognized characters in the images are essential for creating training datasets for the CNN model.

Data annotation tools play a crucial role in the development of machine learning models, especially for tasks like prehistoric Tamil character recognition from stone images. Data annotation involves labeling or marking data to create annotated datasets, which are essential for training and evaluating machine learning models. Here are some data annotation tools that can be used for this specific task:

## Label box:

Description: Labelbox is a versatile platform that allows users to create, manage, and collaborate on data annotation tasks. It supports a wide range of data types, including images, and provides tools for text and object annotation.

## RectLabel:

Description: RectLabel is a macOS application designed for annotating bounding boxes in images. It is suitable for tasks that involve identifying and labeling characters or objects in images.

## VGG Image Annotator (VIA):

Description: VIA is an open-source image annotation tool that is web-based. It supports the annotation of regions (e.g., bounding boxes) and points. It's customizable and can be extended for various annotation tasks.

**Data Management Systems**: Robust data management systems help organize and store the vast amount of image and character data, making it accessible for training and research which includes Data Security Ensuring the security and integrity of the stored data is paramount in a project of this nature. Implementing stringent data security measures, Data Accessibility In addition to managing data, the project should prioritize data accessibility. User-friendly interfaces and well- structured databases enable researchers, historians, and archaeologists to easily access and retrieve the information they need. This accessibility facilitates collaborative research and knowledge sharing. Data Preservation: The project should have mechanisms in place for long-term data preservation. By archiving data in a standardized format, the project ensures that future generations of researchers and scholars will have access to the same data, contributing to the continuity of research efforts and the preservation of historical knowledge.

**Visualization and Reporting Tools:** Software for visualizing the results, generating reports, and sharing findings is crucial for communicating research outcomes. Visualization and reporting tools are essential for analyzing, interpreting, and presenting data in a comprehensible and actionable format. These tools help users create informative charts, graphs, dashboards, and reports, enabling data-driven decision-making and communication. In the context of prehistoric Tamil character recognition from stone images, these tools can be valuable for visualizing results, sharing findings, and supporting research efforts

**Documentation and Reporting Tools:** Tools for documenting research methodologies, findings, and results are essential for academic and research purposes. Documentation and reporting tools are essential components of image processing and computer vision projects, facilitating the organization, documentation, and communication of research findings, analysis results, and project details. These tools are indispensable for conveying information effectively, maintaining a record of project progress.

# CHAPTER 2

**DESIGN AND IMPLEMENTATION**

## PROPOSED SYSTEM Image Acquisition

The system's image acquisition component forms the foundational step in the character recognition process. High-resolution scanners or cameras are employed to capture clear and detailed images of stone inscriptions. These images serve as the primary input data for character recognition. The use of high-quality imaging devices is critical to ensure that the fine details and nuances of the inscribed characters are preserved faithfully. The clarity of these images significantly influences the system's ability to accurately recognize and interpret the prehistoric Tamil characters. Image acquisition is a fundamental step in the field of computer vision, image processing, and various other domains. It involves capturing digital images from the physical world, which can be further processed, analyzed, and utilized for a wide range of applications. The process of image acquisition is achieved through various devices and techniques, and it plays a crucial role in ensuring the quality and reliability of data for subsequent analysis.

## Image Preprocessing

Image preprocessing techniques play a pivotal role in enhancing the quality of the acquired stone inscription images. These techniques include noise reduction, contrast adjustment, and other image enhancements aimed at improving character visibility. By addressing issues such as image artifacts, fading, or irregular lighting, preprocessing ensures that the extracted characters are of the highest quality. This step is crucial in optimizing the performance of the subsequent character recognition process, as it provides a clear and unobstructed canvas for the system to work with. Image preprocessing is a critical phase in the field of

computer vision and image processing, serving as the foundation for successful image analysis and understanding. This process involves a series of operations and techniques applied to raw image data to enhance its quality, reduce noise, correct imperfections, and prepare it for subsequent analysis, such as object detection, recognition, or segmentation. Image preprocessing is essential for improving the accuracy and efficiency of computer vision algorithms.

## Character Extraction

The character extraction component of the proposed system leverages Convolutional Neural Networks (CNNs) for the automated identification and extraction of individual characters from the stone inscription images. Deep learning models are trained to recognize the intricate shapes and forms of prehistoric Tamil characters. This phase of the system is where the true automation takes place, replacing the intensive manual character extraction process with advanced technology. The CNNs are fine-tuned to work with the specific characteristics of prehistoric Tamil scripts, which often feature intricate shapes, loops, and curves.

## Character Classification

Following character extraction, the system employs machine learning techniques to classify the extracted characters into their respective categories within prehistoric Tamil scripts. This classification process is instrumental in organizing the recognized characters and categorizing them accurately. By categorizing characters, the system enables researchers and scholars to navigate and analyze the inscriptions more efficiently, facilitating historical and linguistic interpretation. Character classification is a critical component of the Prehistoric Tamil Character Recognition from Stone Images system, playing a central role in deciphering and understanding ancient scripts inscribed on stone surfaces. This process is essential for historical and archaeological research, as it enables the

identification and interpretation of prehistoric Tamil characters, shedding light on the language, culture, and history of ancient Tamil-speaking civilizations.

The Significance of Character Classification:

Character classification is the process of categorizing individual characters or symbols into predefined classes based on their visual and structural attributes. In the case of prehistoric Tamil character recognition from stone images, this process is vital for several reasons.

## Database Management

Historical Insight: The accurate identification of prehistoric Tamil characters provides invaluable insights into the language, writing system, and historical context of the civilization that inscribed these characters on stone surfaces. It allows researchers and historians to decipher and understand the content of inscriptions.

Cultural Preservation: Character classification contributes to the preservation and documentation of ancient Tamil culture and heritage. It ensures that the knowledge contained in these inscriptions is not lost to time.

Archaeological Understanding: Archaeologists rely on character classification to analyse and date stone inscriptions, contributing to a better understanding of the chronology and historical significance of archaeological sites.

Linguistic Studies: Language scholars and linguists benefit from the accurate classification of prehistoric Tamil characters, enabling them to study the evolution of the Tamil script and its variations over time.

Challenges in Prehistoric Tamil Character Classification,Character classification in the context of prehistoric Tamil scripts comes with its own set of challenges:

Variability: Prehistoric Tamil characters can exhibit significant variability in terms of style, size, and degradation due to weathering and erosion. This variability complicates the task of classification.

Complex Shapes: Some characters in prehistoric Tamil scripts have intricate and complex shapes, including loops, curves, and ornate details. Distinguishing between characters with similar structures is a challenging aspect of classification.

Limited Training Data: Availability of comprehensive training data for prehistoric Tamil characters can be a challenge. The system needs to be trained on a diverse set of characters to handle variations effectively.

Historical and Archaeological Significance: The accuracy of character classification is paramount, as any inaccuracies could lead to misinterpretations of historical and archaeological findings.

## Visualization

The system incorporates visualization tools and reporting mechanisms that allow researchers to interact with and visualize character recognition results. This feature is instrumental in presenting the findings and insights in a clear and comprehensible manner. It enables scholars to generate reports and share their research outcomes effectively. The incorporation of visualization tools and reporting mechanisms in the Prehistoric Tamil Character Recognition from Stone Images system is a pivotal component that plays a fundamental role in the successful interpretation, analysis, and dissemination of research findings. This feature is designed to bridge the gap between raw character recognition results and the meaningful insights that researchers, historians, and scholars seek to derive from prehistoric Tamil inscriptions found on stone images. In the following comprehensive discussion, we will explore the significance, functions, and applications of the visualization and reporting tools within this context.

The significance of visualization tools and reporting mechanisms in the context of prehistoric Tamil character recognition cannot be overstated. These tools serve as a bridge between the complex, algorithm-driven character recognition process and the human comprehension of historical and archaeological findings. By converting recognition results into easily digestible formats, visualization and reporting mechanisms offer several key advantages:

Clarity and Comprehensibility: The visualization of recognized prehistoric Tamil characters and their arrangement in stone inscriptions allows researchers to directly see and comprehend the results. This visual representation aids in the quick and intuitive understanding of the script.

Data Exploration: Visualization tools enable researchers to explore and interact with the recognized characters, providing the flexibility to zoom in, pan across the inscription, and focus on specific segments of interest. This exploration can reveal nuances and patterns in the script.

Data Validation: Researchers can validate the recognition results visually, helping to identify any potential errors or misinterpretations. This visual inspection is critical for ensuring the accuracy of historical findings.

Research Documentation: Visualization tools facilitate the creation of visual documentation for research. Researchers can capture images and representations of recognized characters for inclusion in academic papers, reports, and presentations.

## ARCHITECTURE DESIGN:



Character segmentation

Size normalization

Denoising

Binarization

Input Image

Classification and recognition

Feature extraction

Fig 2.1-Architecture Design

## Input Image

The input image represents the foundation of the recognition process. It contains the visual representation of prehistoric Tamil characters inscribed on stone surfaces. The quality and clarity of this image significantly impact the success of character recognition. For this reason, it's essential to ensure that the input image is of high quality and resolution.

## Binarization

Binarization, following the input image, transforms the visual information into a binary format, converting characters into a distinct foreground against a background. This critical step simplifies subsequent processing and allows the system to focus on the characters, reducing interference from the background. Binarization is a fundamental image processing technique that involves the conversion of a grayscale or colour image into a binary image. In a binary image, each pixel is represented as either black (usually denoted as 0) or white (usually denoted as 1), making it a valuable preprocessing step for various image analysis tasks.

## Denoising

In the denoising phase, the system employs various techniques to remove any artifacts, speckles, or unwanted elements present in the binary image. This step enhances character clarity, which is especially valuable when dealing with inscriptions on stone surfaces, which may exhibit imperfections. Denoising, in the context of image processing, is the process of reducing or removing unwanted noise or disturbances from an image. Noise in an image can manifest as random variations in pixel intensity, which can be caused by factors such as sensor limitations, compression artifacts, or environmental conditions. The goal of denoising is to improve the visual quality of the image and enhance the accuracy of subsequent image analysis tasks.

## Character Segmentation

Character segmentation is a pivotal stage where the individual characters are separated from each other and isolated from the surrounding elements. This separation makes each character a distinct entity, simplifying the recognition process by allowing the system to focus on one character at a time. Character segmentation is a critical image processing task that involves dividing a text image or document into individual character components. This process is

essential in various applications, including optical character recognition (OCR), handwriting recognition, and text analysis. Character segmentation aims to isolate and identify each character, making it easier for subsequent analysis and recognition

## Size Normalization

Size normalization ensures that characters have a uniform size, making them consistent for recognition. This is particularly important when dealing with characters that may vary in size due to factors like inscriptions on stones of different dimensions. Size normalization is a critical step in image processing and analysis, especially when dealing with characters or objects of varying sizes within an image. This process involves making objects or characters in an image uniform in size or scale, which is particularly useful for tasks like character recognition, object detection, and image comparison.

## Feature Extraction

Feature extraction involves the extraction of unique and distinguishing characteristics of each character. Techniques such as Fourier feature mappings and Zernike moments capture valuable information that is used for character recognition, making it easier to differentiate between similar characters. Feature extraction is a crucial process in image processing and computer vision that involves transforming raw image data into a set of meaningful and representative features. These features are designed to capture essential information from the image, making it easier for machine learning algorithms to analyze and recognize patterns.

## Classification and Recognition

In the classification and recognition phase, the system matches the extracted features with known characters in the prehistoric Tamil script. By identifying patterns and similarities, it recognizes characters and provides the output, which may include recognized characters, enabling users to access and utilize the

deciphered information. Classification and recognition are vital processes in image analysis and computer vision that involve identifying and categorizing objects, patterns, or characters within images or data. These processes are crucial for tasks like object recognition, character recognition, and content- based image retrieval. Classification and recognition are foundational processes in computer vision and image analysis, enabling a wide range of applications across various domains. They often rely on machine learning and deep learning techniques to automate the identification and categorization of objects, patterns, or text within images. The accuracy and robustness of these systems depend on the quality of features, the selection of models, and the size and diversity of training data.

## MODULES

The system for prehistoric Tamil character recognition from stone images is composed of several interrelated modules, each serving a distinct role in the character recognition process. These modules work collaboratively to automate the recognition and interpretation of characters inscribed on stone artifacts. Here are the key modules:

## Data Collection Module:

Responsible for gathering a diverse dataset of stone images containing prehistoric Tamil characters. This involves data collection, curation, and organization to ensure a comprehensive dataset. The Data Collection Module is dedicated to assembling a diverse dataset of stone images featuring prehistoric Tamil characters. Its primary responsibilities include sourcing and collecting these images, curating the dataset to ensure its quality and relevance, and organizing the data for effective use. The module plays a pivotal role in providing the foundational material needed for training and testing the prehistoric Tamil character recognition system, making it an essential component in historical and archaeological research applications.

## Binarization Module:

Converts the stone images into binary format, distinguishing characters from the background. This simplifies character extraction and aids in recognition. The Binarization Module is designed to transform stone images into a binary format, effectively separating the characters from the background. By converting the images into a black-and-white representation, it simplifies the process of character extraction, making it more straightforward for subsequent recognition algorithms. This module enhances the clarity of the characters and plays a critical role in improving the overall accuracy and efficiency of the prehistoric Tamil character recognition system.

## Denoising Module:

Focuses on removing unwanted noise and artifacts from the binary images, enhancing image quality and reducing interference. The Denoising Module is specifically designed to eliminate unwanted noise and artifacts from binary images. Its primary goal is to enhance image quality by reducing interference, thereby improving the accuracy and reliability of subsequent image analysis and processing tasks. This module is crucial for ensuring that the preprocessed images are clean and free from any extraneous elements that may hinder the accurate recognition of prehistoric Tamil characters from stone inscriptions.

## Character Segmentation Module:

Isolates individual characters from the processed images, making each character a distinct entity and simplifying the recognition process. The Character Segmentation Module plays a crucial role in the prehistoric Tamil character recognition system. It is responsible for isolating individual characters from the processed stone images, effectively separating each character into distinct entities. This segmentation simplifies the character recognition process, allowing for more accurate analysis and classification of each character. By breaking down the text into individual components, this module contributes to the system's ability

to decipher and understand the prehistoric Tamil script, making it a valuable asset for historical and archaeological research.

## Size Normalization Module:

Ensures that characters are standardized to a consistent size, an important step for accurate recognition. The Size Normalization Module is crucial for standardizing characters within a consistent size, a critical step in ensuring accurate character recognition. By resizing and aligning characters to a uniform scale, it minimizes variations caused by differing character dimensions. This standardization enhances the performance of recognition algorithms, making them more robust to variations in character size and aiding in the decipherment of prehistoric Tamil characters from stone inscriptions.

## Feature Extraction Module:

Extracts relevant character features from the segmented and normalized characters, utilizing techniques like Fourier feature mappings and Zernike moments. The Feature Extraction Module is responsible for capturing significant character attributes from segmented and size-normalized characters obtained from stone images. It employs advanced techniques such as Fourier feature mappings and Zernike moments to extract informative features. These features encode vital details about the characters' shapes, facilitating accurate classification. This module enhances the recognition process by providing valuable data for character differentiation in the Prehistoric Tamil Character Recognition system, contributing to historical and archaeological research applications.

## Model Building Using CNN Module:

Focuses on designing, training, and evaluating a Convolutional Neural Network (CNN) or a similar deep learning model tailored for character recognition. This module includes architecture design, model training, and optimization. The

Model Building Using CNN Module is dedicated to the creation, training, and evaluation of a specialized Convolutional Neural Network (CNN) model tailored for character recognition. It encompasses critical tasks such as architectural design, model training, and optimization, aimed at achieving accurate recognition of prehistoric Tamil characters from stone images. This module is essential for the system's success, as the choice of architecture and effective training are pivotal in ensuring the accuracy and robustness of the character recognition system.

## Classification and Recognition Module:

The central module responsible for character classification and recognition. It matches the extracted features with known characters in the prehistoric Tamil script, providing the recognized characters as the output. The Classification and Recognition Module serves as the core component responsible for character classification and recognition in the prehistoric Tamil script. By leveraging the features extracted from stone images, it matches these features with known characters from the script. The module's primary function is to provide the recognized characters as output, enabling the deciphering and interpretation of ancient inscriptions and contributing to historical and archaeological research.

## MODULES DESCRIPTION Data Collection Module

The Data Collection Module serves as the foundational pillar of the entire recognition system. Its primary purpose is to acquire a diverse and representative dataset of stone images containing prehistoric Tamil characters. The module involves a multifaceted approach to sourcing, curating, and organizing these images to create a comprehensive dataset. The module begins by actively sourcing stone images from various channels. These sources may include historical archives, archaeological collections, fieldwork expeditions to sites with

Tamil inscriptions, and digitization initiatives. The aim is to gather a wide array of stone images representing different time periods, regions, and styles of Tamil script.

Data Curation: Once collected, the stone images undergo a thorough curation process. This involves cataloging the images, annotating their provenance, dating, and relevant historical context. Images are carefully examined for authenticity, and metadata such as location, script variation, and the type of stone are recorded. **Diversity and Representation**: The Data Collection Module is designed to ensure diversity and representation within the dataset. It aims to encompass inscriptions from different historical eras, geographical locations, and styles of prehistoric Tamil characters. This diversity ensures that the recognition system can handle a wide range of characters and historical variations.

**Data Quality**: Quality assurance measures are integrated into the module. This includes checks for image quality, resolution, and data integrity. Any low-quality images or data anomalies are addressed, and, where possible, images are enhanced to improve clarity.

## Denoising Module

The Denoising Module is a critical component within the system for prehistoric Tamil character recognition from stone images. Its primary objective is to enhance the quality of the binary images obtained during the binarization process, ultimately improving the accuracy and reliability of the character recognition process. In the context of historical inscriptions on stone surfaces, it is common to encounter various forms of noise, which can include speckles, blemishes, scratches, and irregularities. These imperfections may arise from factors such as the condition of the stone, the quality of the inscriptions, or the image acquisition process.

The Denoising Module employs a range of sophisticated image processing techniques to mitigate the effects of noise on the binary images. This includes but

is not limited to the application of filters, smoothing algorithms, and morphological operations. The process aims to remove extraneous elements while preserving the integrity of the characters themselves. By systematically eliminating noise, the module significantly improves the clarity and legibility of the characters, making them more amenable to subsequent processing steps. The effectiveness of the Denoising Module is of paramount importance, especially in the context of historical and archaeological research, where the preservation and accurate interpretation of ancient inscriptions are fundamental to understanding cultural heritage and historical narratives. The module ensures that the characters' visual information remains intact while removing any unintended distortions, ultimately contributing to the system's ability to accurately recognize and decipher prehistoric Tamil characters from stone images.

## Character Segmentation Module

Character Segmentation is a pivotal component within the recognition system, serving the crucial function of isolating individual characters from the processed stone images. This module is designed to tackle the complex task of distinguishing characters from the background, as historical inscriptions on stone surfaces may vary in terms of lighting, quality, and the presence of extraneous elements.

The process begins with the input image, which may contain multiple characters inscribed near one another. Character segmentation utilizes various image processing techniques, including but not limited to thresholding, contour detection, and morphological operations. These techniques work in unison to delineate character boundaries and create distinct character regions within the image.

One of the primary challenges in character segmentation is handling the intricacies of prehistoric Tamil script, characterized by unique shapes, loops, and varying character sizes. The module must account for these variations and ensure that characters are accurately separated while minimizing the risk of under-

segmentation (combining characters) or over-segmentation (fragmenting characters).

The segmented characters, once isolated, become the focus of subsequent processing steps, such as size normalization and feature extraction. This accurate segmentation is fundamental for the recognition module to correctly identify and classify individual characters.

To enhance segmentation accuracy, it's often necessary to incorporate adaptive algorithms and heuristics, which adapt to the specific characteristics of the input image. Furthermore, fine-tuning and iterative refinement are common practices to improve segmentation results, ensuring that the system can effectively handle the challenges posed by historical stone inscriptions.

The Character Segmentation Module is an integral part of the system's image processing pipeline, and its successful operation significantly contributes to the overall accuracy and reliability of prehistoric Tamil character recognition from stone images.

## Size Normalization Module

The Size Normalization Module is a critical component in the pipeline for recognizing prehistoric Tamil characters from stone images. It plays a pivotal role in ensuring the uniformity of character dimensions, which is essential for accurate recognition.

In the context of historical stone inscriptions, characters can vary significantly in size due to factors like the type of stone, the period in which the inscriptions were made, and the craftsmanship involved. These variations pose a considerable challenge for character recognition, as algorithms and models may struggle to correctly identify characters that differ in size. The main objective of the Size Normalization Module is to standardize the dimensions of characters, making them consistent and eliminating the influence of size-related variations. This normalization process involves resizing characters to a uniform dimension that is

typically set as a reference size. This reference size can be determined based on historical knowledge or through an analysis of the dataset.

The size normalization process generally follows these steps:

**Character Extraction:** After character segmentation, individual characters are isolated from the stone images. These characters can vary in size and orientation.

**Size Measurement:** The module measures the size of each character. This measurement can be in terms of pixel dimensions or any other suitable unit.

**Resizing:** Characters are then resized to the reference size. This resizing process retains the aspect ratio to prevent distortion and ensures that all characters have the same dimensions.

**Normalization Completion:** After resizing, the characters are now standardized in size, making them uniform and comparable. By standardizing character size, the Size Normalization Module enhances the accuracy of subsequent recognition steps. It ensures that characters of different sizes are processed consistently, reducing the risk of misclassification due to size variations.

## Feature Extraction Module

The Feature Extraction Module is a critical component in the pipeline for recognizing prehistoric Tamil characters from stone images. This module plays a central role in capturing the distinctive characteristics of characters, making them recognizable and distinguishable.

In the context of recognizing prehistoric Tamil characters, feature extraction is a fundamental process that transforms visual information into a format that is suitable for machine learning algorithms to work with. Features are essentially data representations that capture the unique attributes of characters. These attributes can include details like strokes, loops, curves, and other patterns that differentiate one character from another.

The Feature Extraction Module employs various techniques to extract relevant features from the segmented and normalized characters. These techniques are chosen to highlight the most discriminative aspects of the characters. In the context of prehistoric Tamil characters, where there can be significant similarity and variability among characters, effective feature extraction is essential.

Common techniques used in feature extraction for character recognition include:

**Fourier Feature Mappings:** Fourier transforms are applied to character images to represent them in terms of frequency components. This technique is particularly useful for capturing patterns and repeating structures within characters.

**Zernike Moments:** Zernike moments provide a mathematical representation of character shapes and can capture rotational and scale-invariant features. This makes them effective for recognizing characters with varying orientations and sizes.

The choice of feature extraction technique depends on the specific characteristics of the characters being recognized. For prehistoric Tamil characters, the features extracted should emphasize the unique aspects of this script, including the complexity of characters and the presence of intricate shapes, loops, and curves.

Once the features are extracted, they serve as the foundation for character recognition. These feature representations are then used to train machine learning models, such as Convolutional Neural Networks (CNNs), for the task of recognition.

The Feature Extraction Module is a critical bridge between the preprocessing of character images and the subsequent recognition steps. It ensures that the information necessary for recognition is extracted and effectively represented, enabling the system to differentiate between characters and accurately identify them.

In summary, the Feature Extraction Module is a crucial part of the recognition pipeline, responsible for extracting unique features that allow for accurate and reliable prehistoric Tamil character recognition from stone images.

## Classification and Recognition Module

The Classification and Recognition Module is the core component of the recognition system, responsible for the final stage of character recognition in the context of prehistoric Tamil characters inscribed on stone images. This module involves a series of advanced processes to accurately identify and classify characters:

**Feature Matching:** This module utilizes the extracted features from the Feature Extraction step, which include Fourier feature mappings and Zernike moments. These features serve as the distinctive characteristics that allow the system to differentiate between similar characters. The Feature Matching module plays a crucial role in prehistoric Tamil character recognition by leveraging the distinctive features extracted in the previous Feature Extraction step. These features, namely Fourier feature mappings and Zernike moments, serve as unique characteristics that enable the system to distinguish between characters with high similarity and variability.

**Character Classification:** During this phase, the system compares the extracted features with a repository of known prehistoric Tamil characters. It employs machine learning and deep learning techniques to classify characters based on the identified features. The system is trained on a vast dataset of known characters to ensure a high level of accuracy in character classification. Character Classification is a pivotal phase in the prehistoric Tamil character recognition process. During this stage, the system meticulously compares the distinctive features extracted from the stone images with an extensive repository of known prehistoric Tamil characters. It harnesses the power of machine learning and deep learning techniques, leveraging the identified features for precise classification.

The system's training involves exposure to a vast dataset of established characters, equipping it with the knowledge necessary to achieve a high level of accuracy in character classification.

**Recognition Output:** Upon successful character classification, the module provides the recognition output, which includes the identified prehistoric Tamil characters. This output can be in the form of text, digital representations of the characters, or other relevant formats. It enables users, researchers, and historians to access and interpret the inscriptions present in the stone images. The Recognition Output module represents the culmination of the prehistoric Tamil character recognition process. Once character classification is successfully completed, this module delivers the recognized prehistoric Tamil characters to users and researchers. The output is typically presented in various formats, such as plain text, digital representations of the characters, or other relevant forms, making it accessible and understandable.

**Accuracy Assessment:** The module incorporates accuracy assessment mechanisms to evaluate the quality of recognition results. This includes methods to calculate recognition accuracy, identify recognition errors, and ensure the system's reliability in identifying characters, especially in the context of highly similar and variable characters. Accuracy assessment is a critical step in evaluating the performance of image processing and computer vision systems. It involves measuring how well a system or model performs in terms of its ability to correctly classify or recognize objects, patterns, or text within images. Accurate assessment provides insights into the reliability and effectiveness of the system.

**Cultural and Historical Relevance:** In recognizing prehistoric Tamil characters, the module is designed to provide not only character recognition but also information about the historical and cultural context. It can link recognized characters to historical databases and provide additional context and references, contributing to historical and archaeological research.

This module is a critical component of the recognition system, ensuring that prehistoric Tamil characters from stone images are accurately and reliably identified. Its integration of advanced feature matching, classification, and iterative learning mechanisms makes it a robust tool for the preservation and interpretation of historical inscriptions.

## SCREEN SHOTS

The screenshot provided below serves as the input image. This image captures a prehistoric Tamil character meticulously inscribed onto a stone surface

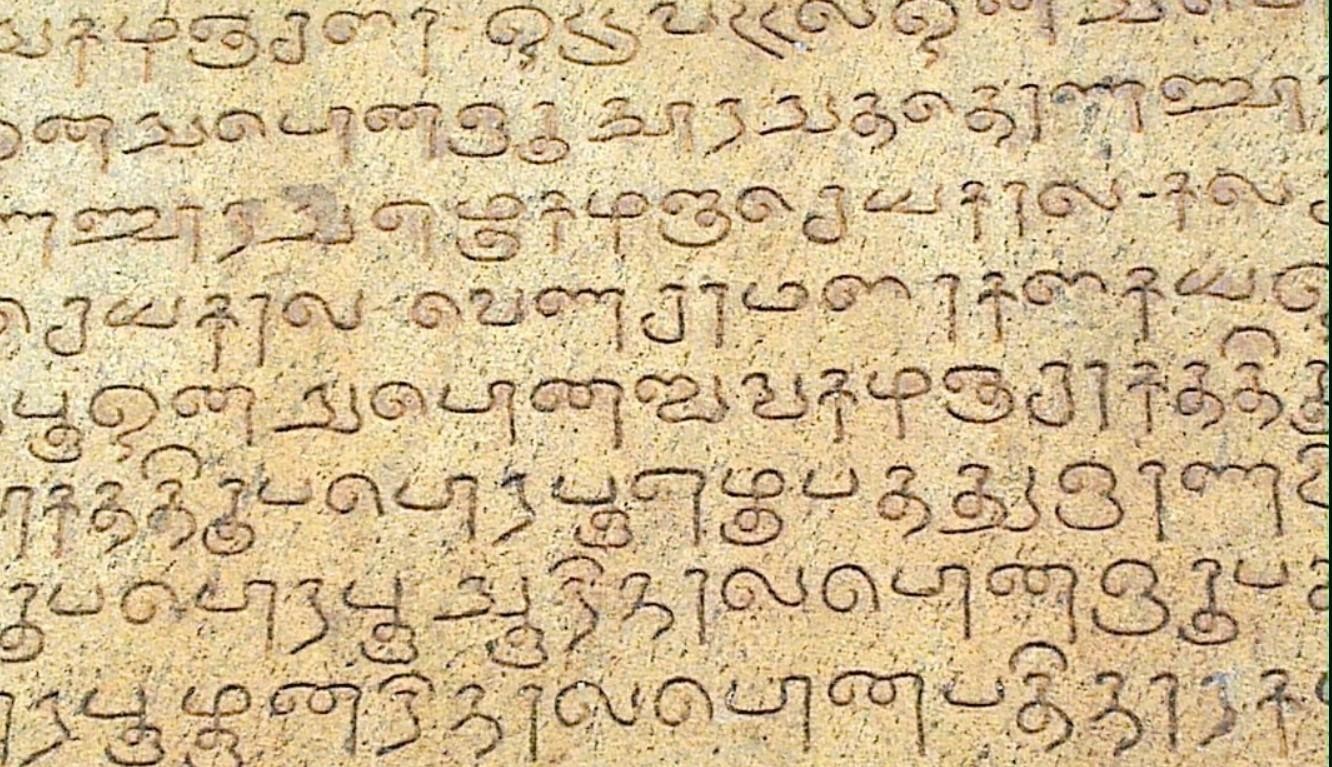


Fig 2.2 – Input Image

The screenshot displayed below represents the denoised image of a prehistoric Tamil character inscribed on a stone artifact, revealing the character's intricate details with enhanced clarity.

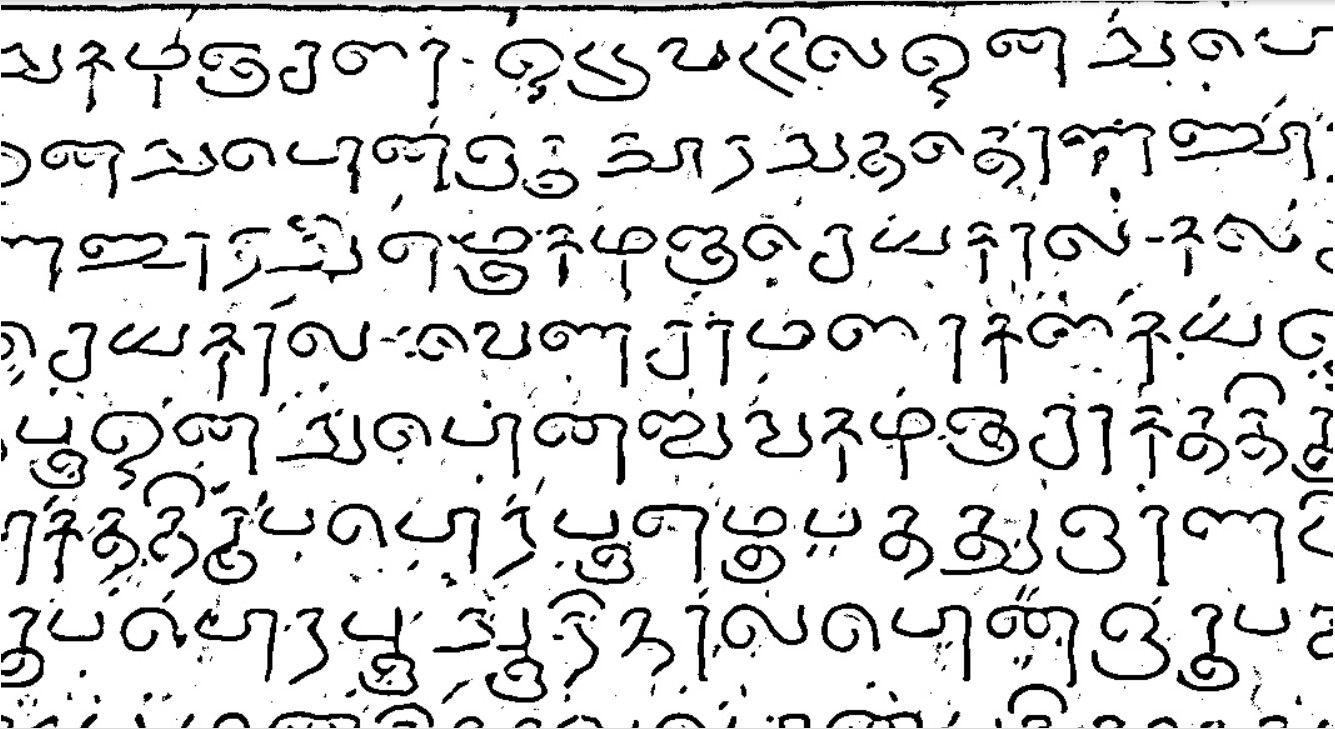


Fig 2.3– Denoised Image

The screenshot provided below showcases the boxed image of a prehistoric Tamil character inscribed on a stone artifact, featuring the character isolated within a bounding box for precise recognition.

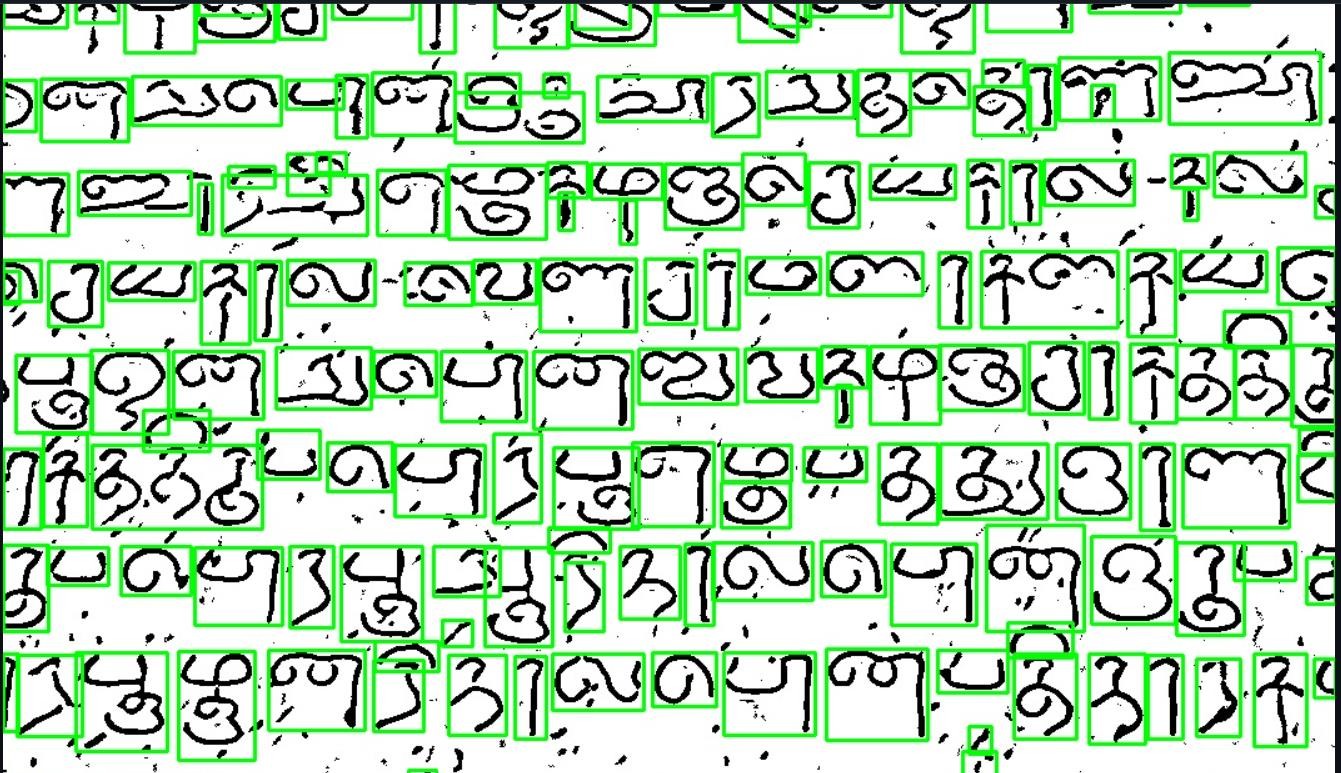


Fig 2.4– Boxed Image

The screenshot provided below showcases the Segmented Character image of a prehistoric Tamil character.

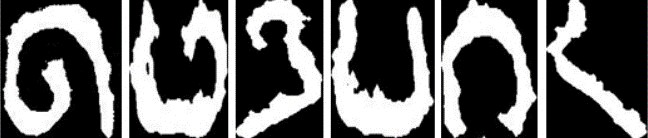


Fig 2.5 – Segmented Character Image

The screenshot below displays the output accuracy image, demonstrating the system's precision in recognizing prehistoric Tamil characters inscribed on stone artifacts.

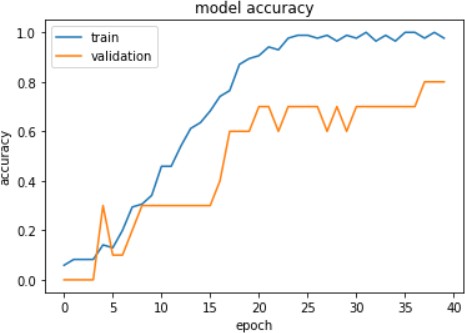


Fig 2.6– Model Accuracy

# CHAPTER 3

**CONCLUSION**

In conclusion, the project focusing on recognizing prehistoric Tamil characters from stone images is a multifaceted endeavour with profound implications for historical research, cultural preservation, and technological innovation. Through the application of cutting-edge technology, including Deep Neural Networks and advanced image processing, the project aspires to significantly improve recognition accuracy, thereby providing a valuable tool for researchers, archaeologists, and historians. The potential to streamline the deciphering of stone inscriptions not only saves time but also enhances the accuracy and efficiency of historical script interpretation. This, in turn, contributes to the preservation and understanding of cultural heritage, protecting invaluable artifacts from potential physical deterioration or loss.The project's outcomes extend beyond the realm of academia, making recognized characters and associated data accessible to researchers and scholars worldwide, fostering collaborative research and knowledge sharing. The innovative use of technology highlights the project's potential to set new standards for character recognition and opens doors to cross-disciplinary collaboration among computer scientists, linguists, archaeologists, and historians. The educational value of the project is notable, offering a resource for students and scholars interested in prehistoric scripts and historical languages, thus contributing to the dissemination of knowledge. With an iterative feedback loop ensuring continual improvement and adaptability to new challenges and datasets, the project represents a dynamic and evolving tool.

It assists experts in the interpretation and transliteration of ancient inscriptions while respecting the essential role of human expertise. Overall, the project embodies a harmonious convergence of technology, cultural preservation, and

historical inquiry, with the potential to enrich our understanding of prehistoric Tamil characters and their significance in the tapestry of histories.

The collaborative nature of the project also promotes intercultural dialogue and understanding. It brings together experts from various disciplines, fostering a spirit of cooperation and mutual learning. As a result, the project not only advances our knowledge of prehistoric Tamil characters but also contributes to the broader goals of multiculturalism and global cooperation.

Finally, the project's adaptability and continual improvement through an iterative feedback loop underscore its commitment to staying at the forefront of technological advancements. This dedication ensures that the project remains relevant and effective in an ever-evolving landscape of historical research and technological innovation.

# REFERENCES

* + 1. Merline Magrina M1 Santhi M2 “Ancient Tamil Character Recognition from Epigraphical Inscriptions using Image Processing Techniques”.

Oct 18-19, 2019, Kings Engineering College, Chennai, TamilNadu, India.

* + 1. Handwritten recognition for Tamil Language: overview, challenges and FutureOn-Line Handwriting Recognition System for Tamil Handwritten Characters -Alejandro H. Toselli1, Mois´es Pastor2, and Enrique Vidal2
    2. Janani. G, Vaishali. V, Mohan (Apr 2016), “Recognition and Analysis of Tamil Inscriptions and Mapping using Image Processing Technique”, IEEE International Conference on Science and Technology Engineering and Management, pp. 181−184
    3. S. RajaKumar, V.Subbiah Bharathi “Eighth century Tamil consonants recognition from stone inscriptions” International conference on Recent Trends in Information Technology(ICRTIT), Pg 40-43, April 2012.
    4. Xiaoqing Ding, Tao Wu “Character Independent Font Recognition on a Single Chinese Character” IEEE Transactions on pattern analysis and machine intelligence, Vol.29, N0.2, pp-195-204, February 2007.
    5. Naveen Kumar Talla and E.V Ramana “Enhancement of South Indian Inscriptions Images Using Denoising and Character Spotting Techniques” International journal of Professional engineering studies, Vol 4, No. 2,

pp-202-207, October 2014.

* + 1. A. A. Prakash and S. Preethi, “Isolated Offline Tamil Handwritten Character Recognition Using Deep Convolutional Neural Network,” International Conference on Intelligent Computing and Communication for Smart World (I2C2SW), pp. 278-281, 2018.
    2. M. V. Jeya Greeba, G.Bhuvaneswari “Recognition of Ancient Tamil character in Stone Inscription using Improved feature extraction” Vol 2, No.3, pp 38-41, February 2014.
    3. R. Prabavathi, J.S. Duela, V. Brindha Devi, S. SenthilPandi, “Prehistoric Stone Image Tamil Character Recognition using Optimized Deep Neural Network using Zernike Moments and Simplex Method,” Turkish J. Comp. Mathem – 2021.
    4. Self-Adaptive Hybridized Lion Optimization Algorithm with Transfer Learning for Ancient Tamil Character Recognition in Stone Inscriptions RD Priya, S Karthikeyan, J Indra- IEEE 2023 <https://tamil.indiatyping.com/index.php/apps/ocr-tamil>

[https://ieeexplore.ieee.org/document/1363950](https://tamil.indiatyping.com/index.php/apps/ocr-tamil)

<https://arxiv.org/abs/1907.04917>

[https://www.academia.edu/48862728/Recognition\_of\_Ancient\_Tamil\_Ch](https://www.academia.edu/48862728/Recognition_of_Ancient_Tamil_Characters_from_Epigraphical_inscriptions_using_Raspberry_Pi_based_Tesseract_OCR) [aracters\_from\_Epigraphical\_inscriptions\_using\_Raspberry\_Pi\_based\_Tes](https://www.academia.edu/48862728/Recognition_of_Ancient_Tamil_Characters_from_Epigraphical_inscriptions_using_Raspberry_Pi_based_Tesseract_OCR) [seract\_OCR](https://www.academia.edu/48862728/Recognition_of_Ancient_Tamil_Characters_from_Epigraphical_inscriptions_using_Raspberry_Pi_based_Tesseract_OCR)

<https://spacy.io/usage/models>