DAYANANDA SAGAR UNIVERSITY

Devarakaggalahalli, Harohalli
Kanakapura Road, Ramanagara - 562112, Karnataka, India

SCHOOL OF ENGINEERING

Bachelor of Technology
in
COMPUTER SCIENCE AND ENGINEERING

Major Project Report

TAMIL SCRIPT TRANSLATION

Batch: 48

By

Maitra M N (ENG20CS0184)
Megha H B (ENG20CS0194)
Pradeep Kumar M (ENG20CS0255)
Rajeshwari N D (ENG20CS0277)

Under the supervision of
Prof. Navya Damodar
Assistant Professor of CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,
SCHOOL OF ENGINEERING
DAYANANDA SAGAR UNIVERSITY,

(2023-2024)

DAYANANDA SAGAR UNIVERSITY

SCHOOL OF ENGINEERING

Department of Computer Science & Engineering

Devarakaggalahalli, Harohalli, Kanakapura Road, Ramanagara - 562112
Karnataka, India

CERTIFICATE

This is to certify that the Major Project Stage-I work titled **"Tamil Script Translation"** is carried out by **Maitra M N (ENG20CS0184), Megha H B(ENG20CS0194), Pradeep Kumar M (ENG20CS0255), Rajeshwari N D(ENG20CS0277),** bonafide students eight semester of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfilment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year **2023-2024.**

Prof. Navya Damodar	**Dr. Girisha G S**	**Dr. Udaya Kumar Reddy K R**
Assistant Professor Dept. of CS&E, School of Engineering Dayananda Sagar University	Chairman CSE School of Engineering Dayananda Sagar University	Dean School of Engineering Dayananda Sagar University
Date:	Date:	Date:

Name of the Examiner **Signature of Examiner**

1.

2.

DECLARATION

We **Maitra M N (ENG20CS0184), Megha H B (ENG20CS0194), Pradeep Kumar M (ENG20CS0255), Rajeshwari N D (ENG20CS0277)** students of eight semester B. Tech in **Computer Science and Engineering**, at School of Engineering, **Dayananda Sagar University**, hereby declare that the Major Project Stage-I titled **"Tamil Script Translation"** has been carried out by us and submitted in partial fulfilment for the award of degree in **Bachelor of Technology in Computer Science and Engineering** during the academic year **2023-2024.**

Student **Signature**

Name1: Pradeep Kumar M
USN : ENG20CS0255

Name2: Megha H B
USN : ENG20CS0194

Name3:Maitra M N
USN : ENG20CS0184

Name4:Rajeshwar ND
USN : ENG20CS0277

Place : Bangalore
Date :

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this project work.

First, we take this opportunity to express our sincere gratitude to School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

We would like to thank ***Dr. Udaya Kumar Reddy K R, Dean, School of Engineering & Technology, Dayananda Sagar University*** *for his constant encouragement and expert advice.*

It is a matter of immense pleasure to express our sincere thanks to ***Dr. Girisha G S Department Chairman****,* ***Computer Science and Engineering****,* ***Dayananda Sagar University,*** *for providing right academic guidance that made our task possible.*

We would like to thank our guide ***Prof. Navya Damodar Assistant Professor****,* ***Dept. of Computer Science and Engineering****,* ***Dayananda Sagar University****, for sparing his valuable time to extend help in every step of our project work, which paved the way for smooth progress and fruitful culmination of the project.*

We would like to thank our ***Project Coordinator Dr. Meenakshi Malhotra*** *and* ***Prof. Mohammed Khurram J*** *as well as all the staff members of Computer Science and Engineering for their support.*

We are also grateful to our family and friends who provided us with every requirement throughout the course.

We would like to thank one and all who directly or indirectly helped us in the Project Work.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
DL	Deep Learning
OCR	Optical Character Recognition
RNN	Recurrent Neural Network
T5	Text-to-Text Transfer Transformer

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The project explores the process of translating Tamil script into English text and subsequently synthesizing English voice output for effective language communication. The translation involves converting Tamil characters into their corresponding English equivalents using computational methods. It utilizes text-to-speech technology to render the translated English text into spoken words, providing a natural and comprehensible voice output. By leveraging computational linguistics and artificial intelligence, this approach facilitates seamless language comprehension and accessibility. The frontend of the project utilizes HTML, CSS, and Brython to create a user-friendly interface for image scanning and uploading, while the backend, powered by Flask, processes uploaded images for accurate translation.Testing reveals the project's accuracy, which is 91%. Recognized as a valuable resource for improvement, it enhances cross-language understanding and promotes more inclusive communication among people speaking different languages.This project is recognized as a valuable tool for promoting cross-language understanding and fostering inclusive communication among individuals who speak different languages. Its implementation of computational linguistics and artificial intelligence contributes to seamless language comprehension and accessibility, making it a significant resource for enhancing communication in multilingual contexts.

CHAPTER 1 INTRODUCTION

1.1 Overview

The Tamil Translation Project is a comprehensive initiative designed to enhance communication across diverse linguistic communities. Recognizing language as a vital channel for cultural exchange and mutual understanding, this project aims to facilitate seamless access to content for Tamil-speaking audiences. The project adopts a meticulous workflow, starting with gathering essential linguistic resources and gaining a deep understanding of the context and intricacies of the source material. Through dedicated translation efforts, the project aims to uphold the integrity and purpose of the original content while ensuring linguistic precision and cultural relevance in the Tamil rendition.

A key focus of the Tamil Translation Project is a commitment to cultural sensitivity and adaptation. Translators navigate linguistic and cultural nuances skillfully, employing strategies to effectively convey the intended message to Tamil-speaking audiences. This involves not only linguistic translation but also adapting the content to align with cultural norms and preferences. By embracing cultural diversity and promoting inclusivity, the project aims to create translations that resonate authentically with Tamil-speaking communities, fostering a sense of connection and belonging.

The Tamil Translation Project emphasizes collaboration, involving stakeholders and native Tamil speakers throughout the translation process. Feedback and insights from these key stakeholders inform iterative revisions, ensuring the quality and effectiveness of the translated material. Rigorous testing and validation further uphold standards of excellence, delivering translated content that meets the diverse needs and expectations of audiences. Ultimately, the Tamil Translation Project demonstrates the transformative impact of language, fostering mutual understanding and enriching cross-cultural dialogue in our interconnected world.

1.2 SCOPE AND SIGNIFICANCE OF THIS PROJECT

The Tamil Translation Project covers a wide array of content types and subjects to meet the varied needs and interests of Tamil-speaking audiences. It encompasses literature, academic texts, marketing materials, and technical documentation, offering comprehensive translation services across different domains. The project also includes the translation of multimedia content like videos, audio recordings, and digital media to ensure a seamless experience across various mediums. Additionally, it extends to both traditional and modern platforms, including websites, mobile apps, and social media channels, aiming to engage audiences across digital landscapes. By providing a diverse range of translation services, the project aims to meet the linguistic and cultural needs of Tamil-speaking communities worldwide, enabling access to information and fostering cross-cultural exchange.

1. Social Impact
   - Cultural Preservation
   - Educational Empowerment
   - Community Participation

2. Environmental Impact
   - Conservation of Cultural Heritage
   - Emphasis on Sustainable Practices

3. Technical Impact
   - Advancements in Linguistic Studies
   - Exploration of Digital Humanities
   - Promotion of Interdisciplinary Collaboration

CHAPTER 2 PROBLEM DEFNITION

The Tamil Translation Project tackles several significant challenges inherent in translating content from Tamil into English while upholding linguistic accuracy, cultural relevance, and accessibility. One key challenge stems from the intricacies of the Tamil language itself, including its rich history, diverse dialects, and unique script. Translators must navigate these complexities to ensure that the translated content resonates authentically with English-speaking audiences from various regions and cultural backgrounds. Additionally, the project faces the task of preserving the original intent and meaning of the source material while adapting it to suit English language nuances and cultural contexts. This demands careful consideration of context, tone, and cultural references to effectively convey the intended message. Translators must stay updated on new terms, expressions, and concepts across different fields to ensure accurate and contemporary translations. Adapting content to align with English cultural norms while respecting the integrity of the original material requires nuanced cultural understanding and sensitivity.

Consistency and quality are also critical challenges, especially when dealing with large volumes of content or multiple translators. Establishing clear guidelines, standards, and quality assurance processes is vital to maintaining the overall quality and coherence of the translated material. Furthermore, the project must prioritize accessibility, ensuring that translated content is available and understandable to diverse audiences, including those with varying levels of literacy or technological proficiency. By addressing these challenges, the Tamil Translation Project aims to deliver high-quality translations that bridge linguistic and cultural divides, promoting greater understanding and inclusivity within English-speaking communities and beyond.

CHAPTER 3 LITERATURE REVIEW

The literature review for the Tamil Translation Project surveys existing research and practices in translation studies, linguistic anthropology, and cultural communication, focusing on the complexities involved in translating content from Tamil into English. Studies delve into the intricacies of language translation, emphasizing the critical aspects of linguistic accuracy, cultural sensitivity, and audience engagement for producing effective translations. Additionally, scholarly works explore the impact of technology on translation processes, including machine translation tools and localization software, offering innovative approaches to enhance translation efficiency and quality. Moreover, research on cultural adaptation and cross-cultural communication informs the project's strategies for navigating cultural nuances and ensuring the authenticity and relevance of translated content for English-speaking audiences. By synthesizing insights from diverse scholarly perspectives, the literature review establishes a theoretical framework and practical guidance to inform the methodology and approach of the Tamil Translation Project.

Author's Name/ Paper Title	**Conference/Journal Name and year**	**Technology/ Design**	**Results shared by author**	**What you infer**
Ishwari S. Kulkarni/ PROPOSED DESIGN TO RECOGNIZE ANCIENT SANSKRIT MANUSCRIPTS WITH TRANSLATION	Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2022	OCR(Optical Character Recognition) Image Processing	The purpose of this system is to provide access to ancient knowledge.	The research paper outlines a novel approach for the recognition and translation of ancient Sanskrit manuscripts, facilitating accessibility and understanding of their content.
Shikha Chadha/ Ancient Text Character Recognition Using Deep Learning	International Journal of Engineering Research and Technology (2020)	CAE (Convolutional Auto Encoder)	The author Shikha Chadha shared results on ancient text character recognition using deep learning.	Shikha Chadha presented outcomes demonstrating the utilization of deep learning for recognizing characters in ancient texts.

Table 1: Literature survey 1

Author's Name/ Paper Title	Conference/Journal Name and year	Technology/ Design	Results shared by author	What you infer
Ishwari S. Kulkarni/ PROPOSED DESIGN TO RECOGNIZE ANCIENT SANSKRIT MANUSCRIPTS WITH TRANSLATION	Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2022	OCR(Optical Character Recognition) Image Processing	The purpose of this system is to provide access to ancient knowledge.	The research paper outlines a novel approach for the recognition and translation of ancient Sanskrit manuscripts, facilitating accessibility and understanding of their content.
Shikha Chadha/ Ancient Text Character Recognition Using Deep Learning	International Journal of Engineering Research and Technology (2020)	CAE (Convolutional Auto Encoder)	The author Shikha Chadha shared results on ancient text character recognition using deep learning.	Shikha Chadha presented outcomes demonstrating the utilization of deep learning for recognizing characters in ancient texts.

Table 2: Literature survey 2

CHAPTER 4 PROJECT DESCRIPTION

The Tamil-to-English Translation Project aims to enhance communication and cultural exchange by translating a diverse range of content into English, a widely spoken language globally. Recognizing the importance of linguistic accessibility and cultural sensitivity, this project endeavors to provide high-quality translations that resonate authentically with English-speaking audiences. The project commences with establishing a robust translation framework, including acquiring essential linguistic resources like dictionaries, grammar guides, and style manuals. Translators undergo thorough training to ensure proficiency in both the source language and English, enabling them to accurately convey the meaning and intent of the original content while adapting it to English language nuances and cultural contexts.

The content for translation covers various domains such as literature, academic research, marketing materials, technical documentation, and multimedia content. Each translation undergoes rigorous review and quality assurance processes to ensure accuracy, consistency, and cultural relevance. Advanced technology and tools for contextual analysis, terminology management, and quality control are utilized to streamline the translation process and improve efficiency. Collaboration with stakeholders and feedback from native English speakers are essential for refining translations and ensuring their effectiveness. Throughout the project lifecycle, regular communication and evaluation mechanisms are employed to monitor progress, address challenges, and incorporate insights for continuous enhancement.

The primary objective of the Tamil-to-English Translation Project is to promote linguistic inclusivity, cross-cultural understanding, and knowledge dissemination within English-speaking communities and beyond. By delivering high-quality translations that cater to the diverse needs and interests of English audiences, this project contributes to enriching the linguistic landscape and fostering greater global connectivity.

Tolkāppiyam consonants		Vowels											
		ø a	ா ā	ி i	ீ ī	ு u	ூ ū	ெ e	ே ē	ை ai	ொ o	ோ ō	ௌ au
ø (Independent)		அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ
க்	k	க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ	கோ	கௌ
ங்	ṅ	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙை	ஙொ	ஙோ	ஙௌ
ச்	c	ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ	சோ	சௌ
ஞ்	ñ	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ	ஞோ	ஞௌ
ட்	ṭ	ட	டா	டி	டீ	டு	டூ	டெ	டே	டை	டொ	டோ	டௌ
ண்	ṇ	ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணை	ணொ	ணோ	ணௌ
த்	t	த	தா	தி	தீ	து	தூ	தெ	தே	தை	தொ	தோ	தௌ
ந்	n	ந	நா	நி	நீ	நு	நூ	நெ	நே	நை	நொ	நோ	நௌ
ப்	p	ப	பா	பி	பீ	பு	பூ	பெ	பே	பை	பொ	போ	பௌ
ம்	m	ம	மா	மி	மீ	மு	மூ	மெ	மே	மை	மொ	மோ	மௌ
ய்	y	ய	யா	யி	யீ	யு	யூ	யெ	யே	யை	யொ	யோ	யௌ
ர்	r	ர	ரா	ரி	ரீ	ரு	ரூ	ரெ	ரே	ரை	ரொ	ரோ	ரௌ
ல்	l	ல	லா	லி	லீ	லு	லூ	லெ	லே	லை	லொ	லோ	லௌ
வ்	v	வ	வா	வி	வீ	வு	வூ	வெ	வே	வை	வொ	வோ	வௌ
ழ்	ḻ	ழ	ழா	ழி	ழீ	ழு	ழூ	ழெ	ழே	ழை	ழொ	ழோ	ழௌ
ள்	ḷ	ள	ளா	ளி	ளீ	ளு	ளூ	ளெ	ளே	ளை	ளொ	ளோ	ளௌ
ற்	ṟ	ற	றா	றி	றீ	று	றூ	றெ	றே	றை	றொ	றோ	றௌ
ன்	ṉ	ன	னா	னி	னீ	னு	னூ	னெ	னே	னை	னொ	னோ	னௌ

Grantha compound table

Fig 4.1 Compound Table of Tamil letters

4.1 SYSTEM DESIGN

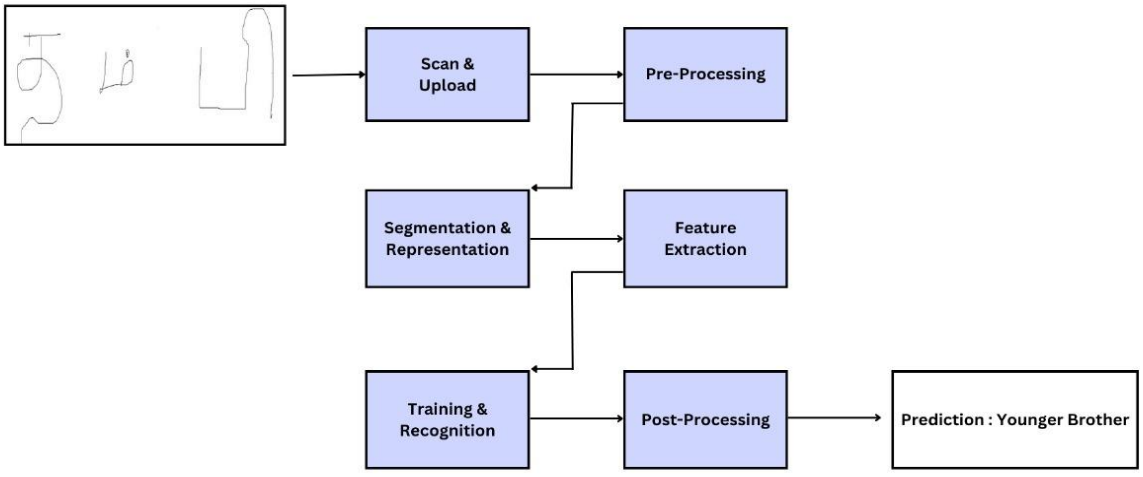


Fig 4.1.1 Block Diagram

4.2 ASSUMPTION AND DEPENDENCIES

4.2.1 Assumption:

1. Recognizing that Tamil and English belong to different language families (Dravidian and Indo-European, respectively), with distinct grammar, syntax, and vocabulary structures.

2. Acknowledging the need to transliterate Tamil characters into English script, considering variations in phonetics and the absence of direct one-to-one mappings between Tamil and English letters.

3. Understanding that translation involves more than linguistic conversion; it requires cultural adaptation to ensure that the meaning and intent of the original text are preserved and effectively conveyed in English.

4. Considering the target audience's familiarity with Tamil culture and language, as well as their proficiency in English, to tailor translations accordingly.

5. Leveraging translation tools and software for efficiency, while acknowledging their limitations in accurately capturing context, idiomatic expressions, and cultural nuances.

6. Implementing rigorous review processes by proficient bilingual translators to ensure accuracy, consistency, and cultural relevance in the translated content.

7. Recognizing that translation is an iterative process, subject to ongoing refinement based on feedback, linguistic advancements, and evolving cultural norms.

These assumptions guide the translation process and inform strategies for effectively bridging linguistic and cultural gaps between Tamil and English**.**

4.2.2 Dependencies:

1. Proficiency in both Tamil and English languages is essential for accurate translation. Translators must understand the grammar, syntax, semantics, and idiosyncrasies of both languages.

2. Dependence on transliteration tools or methods to convert Tamil characters into corresponding English letters or phonetic equivalents. This process requires mapping Tamil sounds to English sounds.

3. Access to comprehensive dictionaries, glossaries, and terminology databases to facilitate accurate translation of specific terms, technical jargon, and domain-specific language.

4. Ability to interpret the context of the original text and convey its intended meaning effectively in English, considering cultural references, idiomatic expressions, and rhetorical devices.

5. Utilization of machine translation (MT) tools, translation memory (TM) systems, and computer-assisted translation (CAT) software to aid in the translation process, improve efficiency, and ensure consistency.

6. Incorporation of feedback from stakeholders, native speakers, and domain experts to refine translations iteratively and address linguistic nuances or cultural sensitivities.

7. Awareness of ethical issues in translation, such as maintaining confidentiality, respecting cultural differences, and avoiding biases or stereotypes.

These dependencies highlight the multifaceted nature of Tamil script translation to English and emphasize the importance of linguistic competence, cultural sensitivity, technological support, and quality assurance in achieving accurate and effective translations.

CHAPTER 5 REQUIREMENTS

5.1. Functional Requirements:

- The system should accurately translate Tamil script into English text, maintaining the meaning and context of the original content.
- Incorporate NLP techniques to handle complex sentence structures, idiomatic expressions, and linguistic nuances during translation.
- Implement voice synthesis capabilities to convert translated English text into spoken words with natural intonation and clarity.
- Enable translation of diverse content types including literature, technical documents, multimedia, and online content.
- Ensure translations are contextually appropriate and culturally sensitive to resonate with English-speaking audiences.
- Handle large volumes of translation requests efficiently, maintaining performance and response times.

5.2. Non Functional Requirements:

- Deliver accurate and reliable translations to avoid misinterpretations or errors in the converted text and voice output.
- Provide a user-friendly interface for translators and end-users, making the translation and voice synthesis process intuitive and accessible.
- Ensure data privacy and confidentiality of translated content, especially for sensitive information.
- Allow for customization and configuration options to adapt translation outputs based on user preferences and domain-specific requirements.
- Implement high-quality voice synthesis algorithms to generate clear, natural-sounding English voice from translated text.
- Optimize system performance to handle real-time voice synthesis efficiently, minimizing latency and resource utilization.

5.3. System/Software Requirements

- Programming languages: Python
- Deep Learning frameworks: TensorFlow, keras
- Image Processing libraries: opencv
- Model Evaluation metric: Confusion matrix, accuracy
- Visual Studio Code
- Modern multi core Processor (e.g., Intel Core i5 or higher) for efficient deep learning computations
- Sufficient RAM (8 GB or more) to accommodate data processing and model training
- Adequate SSD for storing images, datasets and models

CHAPTER 6 METHODOLOGY

1. Data Collection:

Data collection for the Tamil script translation project involves gathering diverse datasets of written Tamil text, parallel corpora aligning Tamil and English sentences, linguistic resources such as dictionaries and grammatical rules, and cultural references specific to Tamil language and context. Additionally, audio recordings of native English speakers are collected to build a dataset for training voice synthesis models. Domain-specific content and user feedback are also important aspects of data collection to enhance translation accuracy and voice synthesis quality. Ethical considerations guide the collection of voice data and ensure compliance with privacy regulations. Continuous updates and expansions of datasets are essential to adapt to evolving language variations and user needs.

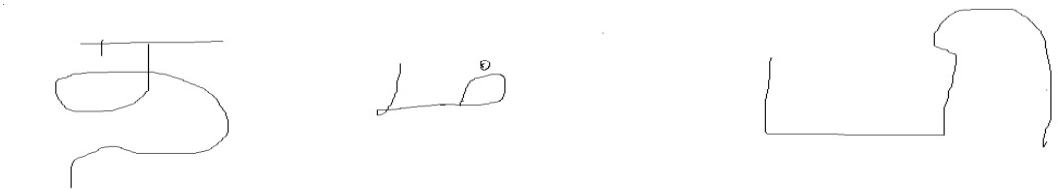


Fig 6.1 Original Image

2. Preprocessing:

In the preprocessing phase of the Tamil script translation project, data undergoes meticulous cleaning and formatting to prepare it for subsequent processing. This involves removing extraneous characters, such as special symbols or HTML tags, to ensure that the text is clean and uniform. Tokenization is then applied to segment the cleaned text into individual words or tokens, facilitating further analysis. Additionally, normalization techniques standardize the text by resolving variations in word forms and representations. Stopword removal helps filter out common words that do not contribute significant meaning to the text. Furthermore, language-specific preprocessing steps are implemented to address unique linguistic features of Tamil, such as handling compound words and inflections. This careful preprocessing ensures that the data is ready for effective machine

translation and voice synthesis, enabling accurate and culturally appropriate translations from Tamil -to-English.

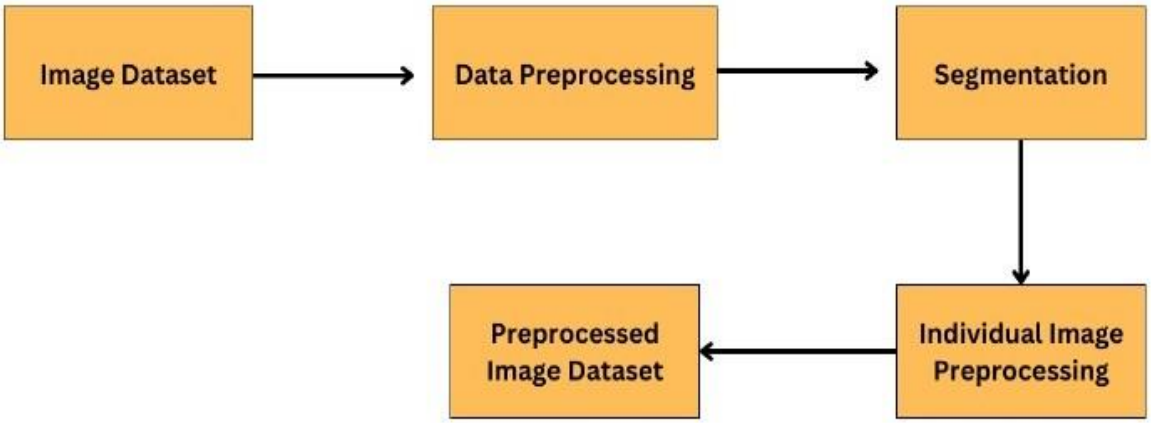


Fig 6.2 Preprocessing Diagram

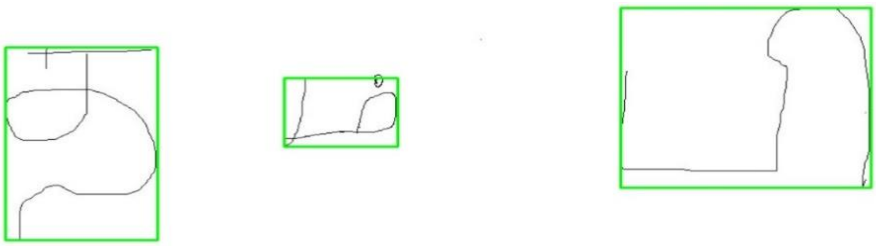


Fig 6.3 Segmented Image

3. Training/Modelling:

In the modelling and training phase of the Tamil script translation project, machine learning models are developed and trained using the preprocessed data to facilitate accurate translation from Tamil to English. This involves selecting appropriate algorithms, such as sequence-to-sequence models based on recurrent neural networks (RNNs) or transformer models like BERT (Bidirectional Encoder Representations from Transformers). The selected models are trained on aligned pairs of Tamil and English sentences from the prepared datasets, optimizing for translation quality metrics like BLEU (Bilingual Evaluation Understudy) score.

During training, the models learn to map input Tamil sequences to corresponding output English sequences, capturing semantic and syntactic relationships between languages. Training involves iterative optimization of model parameters using gradient-based optimization techniques like stochastic gradient descent (SGD) or Adam optimizer. Hyperparameters such as learning rate, batch size, and model architecture are tuned to improve performance and convergence.

Validation data is used to monitor model performance during training, preventing overfitting and ensuring generalization to unseen data. Techniques like early stopping and model checkpointing are employed to save the best-performing model weights.

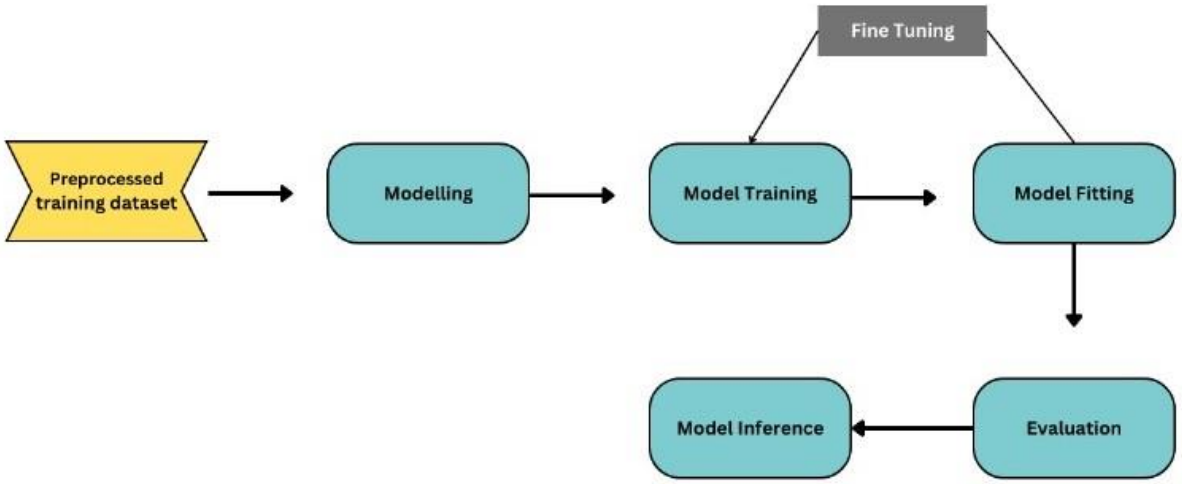


Fig 6.4 Training a module Diagram

4. Evaluation/Inference Model:

In the evaluation and inference phase of the Tamil script translation project, the trained machine translation model is rigorously assessed for its performance and effectiveness in translating Tamil text to English. This evaluation involves several key steps to measure the quality and accuracy of the translation output. First, the model is evaluated using a separate test dataset that was not used during training or validation. This helps assess the model's ability to generalize to unseen data and detect any potential overfitting issues.

Evaluation metrics such as BLEU (Bilingual Evaluation Understudy) score, METEOR (Metric for Evaluation of Translation with Explicit Ordering), and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) are calculated to quantify the similarity between the model-generated translations and human reference translations. These metrics provide valuable insights into the translation quality, fluency, and adequacy of the model.

In real-world inference scenarios, the trained model is deployed to translate Tamil text into English on-demand. The inference process involves feeding input Tamil sequences into the model and generating corresponding English translations as output. The translated text is then further processed, if necessary, for post-editing or integration with voice synthesis systems to produce spoken English output.

By leveraging robust evaluation techniques and efficient inference procedures, the project aims to deliver high-quality translation services that bridge linguistic barriers and foster cross-cultural communication.

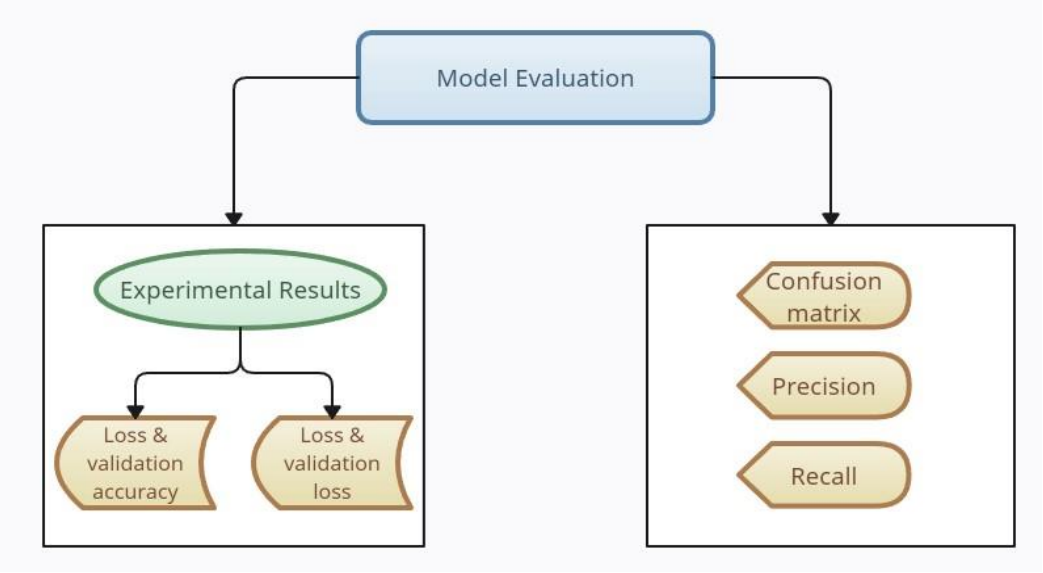


Fig 6.5 Evaluation/Inference mode

CHAPTER 7 EXPERIMENTATION

7.1 SOFTWARE DEVELOPMENT

Our project aims to develop a software application that is capable of translating Tamil script images into English text. Leveraging HTML, CSS, Brython, and Flask, we have created a user-friendly frontend interface for scanning and uploading images. The backend, powered by Flask, processes the uploaded images using advanced algorithms to accurately translate Tamil script into English.

Requirement Analysis: We conducted extensive research to understand the requirements and challenges of translating Tamil script into English text from images.

Frontend Development: Using HTML and CSS, we designed an intuitive and user-friendly interface for scanning and uploading images.

Integration of Brython: Brython a Python 3 implementation for the web, was integrated into the frontend to facilitate seamless communication with the backend.

Backend Development with Flask: Flask, a micro web framework, was utilized to develop the backend logic for image processing and translation.

Image Processing: We employed advanced image processing techniques to preprocess the scanned images, ensuring optimal translation accuracy.

Translation Algorithm: Our software incorporates a robust translation algorithm specifically tailored for accurately translating Tamil script into English text.

Testing and Quality Assurance: Rigorous testing procedures were implemented to validate the accuracy and performance of the translation software.

Deployment and Maintenance: Upon successful testing, the software was deployed to ensure accessibility to users. Regular maintenance and updates are planned to enhance functionality and address any issues that may arise.

By adopting a systematic approach to software development and leveraging cutting-edge technologies, we have successfully created a powerful tool for translating Tamil script images into English text, catering to the diverse needs of our users and quality assurance ensures that the software meets the highest standards of performance and usability.

7.2 CODE SNIPPET

tamil-recognition-training-part-main > translate.py > translate_tamil

```python
from selenium.webdriver import Chrome,ChromeOptions,ChromeService
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
import os
from config import TRANSLATE
import time


def translate_tamil(tamil_word):
    chromedriver_path=os.getcwd()+'\\chromedriver.exe'

    service=ChromeService(executable_path=chromedriver_path)
    options=ChromeOptions()
    # options.add_argument('--headless')


    driver=Chrome(service=service,options=options)

    driver.get(TRANSLATE)

    input_box=driver.find_elements(By.TAG_NAME,"textarea")
    input_box=input_box[1]
    input_box.send_keys(tamil_word)

    time.sleep(3)

    output=driver.find_elements(By.XPATH,"//span[@class='Y2IQFc']")
    output=output[2].text

    time.sleep(10)

    return output
```

CHAPTER 8 RESULTS AND DISCUSSION

Our project aims to create a software application that simplifies the process of translating Tamil text into English. Here's how it works Users upload images containing Tamil text to the software interface, The software scans the uploaded images to identify and extract the Tamil text, Once the Tamil text is extracted, the software translates it into English using a built-in translation mechanism. The translated text is then presented to the user in an easily understandable format, allowing them to comprehend the content without needing knowledge of Tamil.

By automating the translation process from Tamil to English, your software provides users with a convenient way to understand and interpret Tamil text, opening up communication and access to information across language barriers.

Using HTML and CSS, we designed an intuitive and user-friendly interface for scanning and uploading images. Flask, a micro web framework, was utilized to develop the backend logic for image processing and translation.

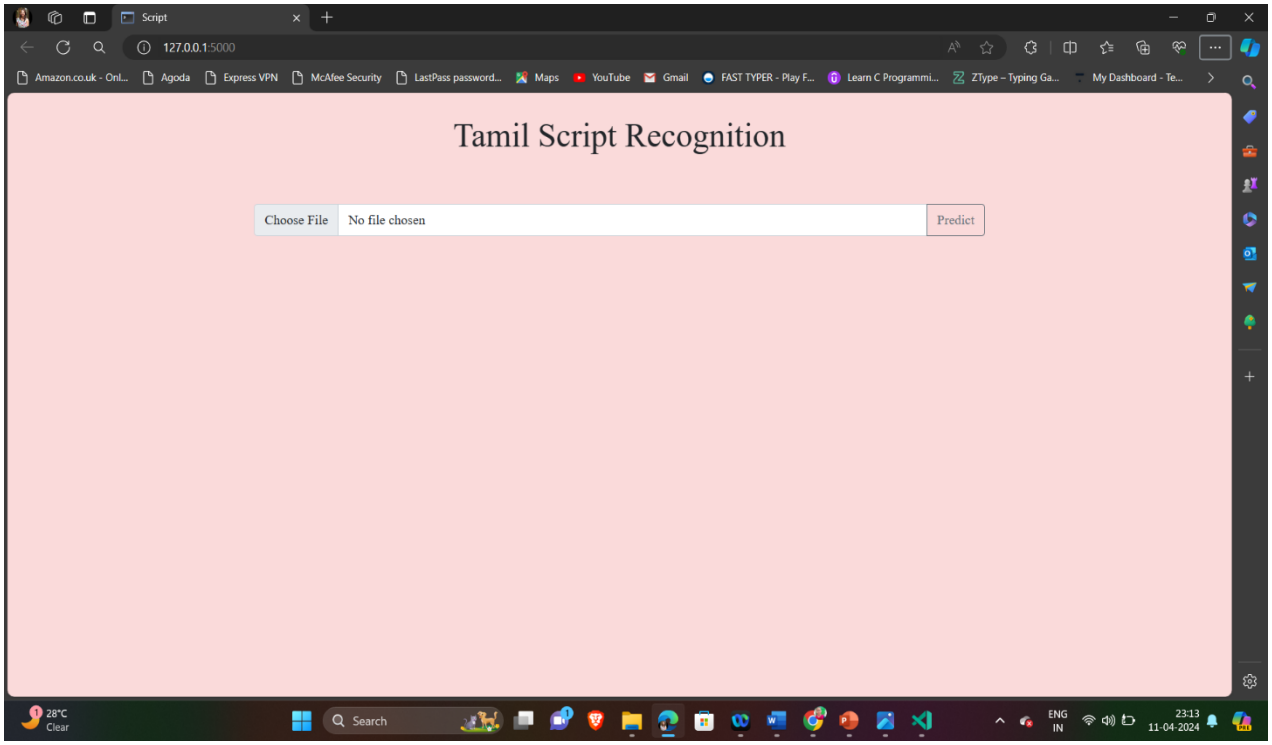


Fig 8.1 User Interface

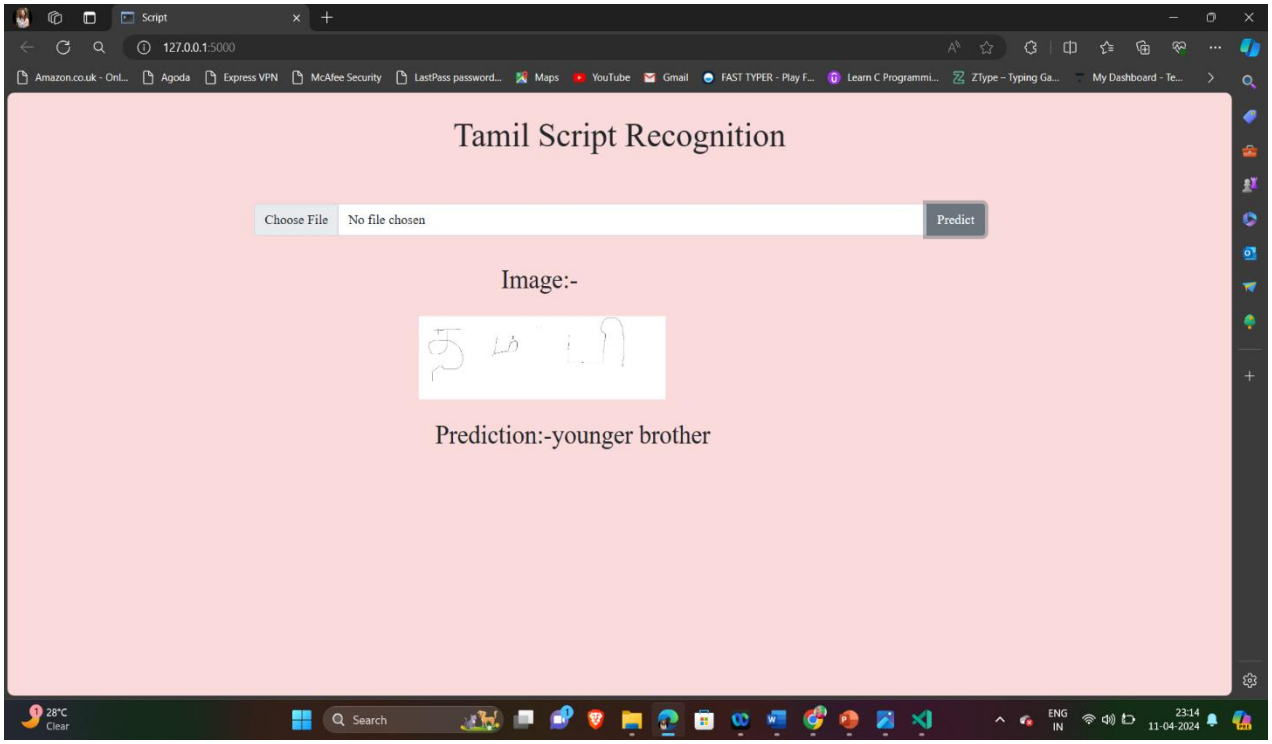


Fig 8.2 Predicted Interface

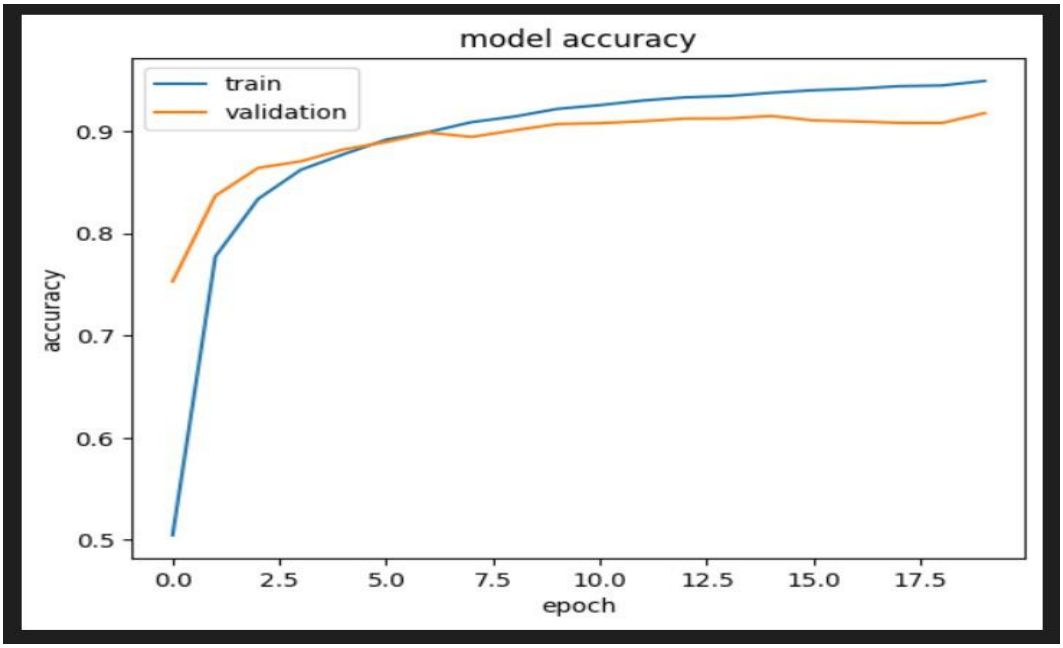


Fig 8.3 Training and Validation Accuracy

```
history = model.fit(X, y, batch_size=32, epochs=10, validation_split=0.1)
```

```
Epoch 6/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 60s 42ms/step - accuracy: 0.8932 - loss: 0.3152 - val_accuracy: 0.8895 - val_loss: 0.3467
Epoch 7/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 61s 43ms/step - accuracy: 0.9025 - loss: 0.2869 - val_accuracy: 0.8986 - val_loss: 0.3265
Epoch 8/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 61s 43ms/step - accuracy: 0.9105 - loss: 0.2640 - val_accuracy: 0.8945 - val_loss: 0.3413
Epoch 9/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 62s 43ms/step - accuracy: 0.9160 - loss: 0.2426 - val_accuracy: 0.9010 - val_loss: 0.3117
Epoch 10/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 61s 43ms/step - accuracy: 0.9239 - loss: 0.2201 - val_accuracy: 0.9071 - val_loss: 0.3038
Epoch 11/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 62s 43ms/step - accuracy: 0.9284 - loss: 0.2087 - val_accuracy: 0.9079 - val_loss: 0.3009
Epoch 12/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 81s 42ms/step - accuracy: 0.9328 - loss: 0.1959 - val_accuracy: 0.9098 - val_loss: 0.3046
Epoch 13/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 59s 42ms/step - accuracy: 0.9351 - loss: 0.1854 - val_accuracy: 0.9124 - val_loss: 0.3105
Epoch 14/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 60s 42ms/step - accuracy: 0.9356 - loss: 0.1821 - val_accuracy: 0.9126 - val_loss: 0.2939
Epoch 15/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 60s 42ms/step - accuracy: 0.9414 - loss: 0.1686 - val_accuracy: 0.9150 - val_loss: 0.2975
Epoch 16/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 61s 43ms/step - accuracy: 0.9426 - loss: 0.1620 - val_accuracy: 0.9106 - val_loss: 0.2929
Epoch 17/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 61s 43ms/step - accuracy: 0.9437 - loss: 0.1547 - val_accuracy: 0.9096 - val_loss: 0.3124
Epoch 18/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 61s 43ms/step - accuracy: 0.9459 - loss: 0.1514 - val_accuracy: 0.9083 - val_loss: 0.3320
Epoch 19/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 61s 42ms/step - accuracy: 0.9491 - loss: 0.1458 - val_accuracy: 0.9081 - val_loss: 0.3222
Epoch 20/20
1426/1426 ━━━━━━━━━━━━━━━━━━━━ 60s 42ms/step - accuracy: 0.9501 - loss: 0.1385 - val_accuracy: 0.9177 - val_loss: 0.2988
```

Fig 8.4 Accuracy and Loss of few epochs of Baseline CNN model

CHAPTER 9 CONCLUSION AND FUTURE WORK

9.1 CONCLUSION

In summary, project has successfully developed a tool that translates Tamil words into English and even speaks them out loud. The project aims to develop user-friendly web application that allows users to upload images containing tamil script. The software processes the images, identifies the script, and provides translated text and appropriate voice as output. Through rigorous testing, we've confirmed its accuracy as 91%. Moreover, the system's translations are highly accurate, offering invaluable insights into the language, culture, history, and context of the civilization that utilized the script. To further aid comprehension and utilization, the project plans to develop educational materials, tutorials, and user guidelines.

9.2 SCOPE FOR FUTURE WORK

In the future, the project aims to expand its capabilities beyond single-word translations to include entire sentences. This means that instead of only understanding individual words, system will be able to understand and translate complete thoughts and ideas. This will be really helpful because people often communicate using full sentences, and being able to translate them accurately will make it much easier for everyone to understand each other, even if they speak different languages. To achieve this goal, the project will need to enhance its translation algorithms and employ more technology to analyze and translate entire sentences while preserving their intended meaning. Additionally, collecting a larger dataset of sentence pairs in both languages will be necessary to train the translation models effectively. Furthermore, updates to the user interface will be required to facilitate the input and display of full sentences, ensuring that the project remains user-friendly and accessible to all users

CHAPTER 10 REFERENCES

[1] Ancient Text Character Recognition Using Deep Learning. International Journal of Engineering Research and Technology(2020)

[2] Ancient Marathi Script Recognition from Stone Inscriptions, Bapu Chendage, Rajivkumar Mente, 5th International Conference On "Innovations in IT and Management".

[3] Recognition of Ancient Tamil Brahmi Characters from Epigraphy. A.Vidhyavani,Journal of Pharmaceutical Negative Results (2022).

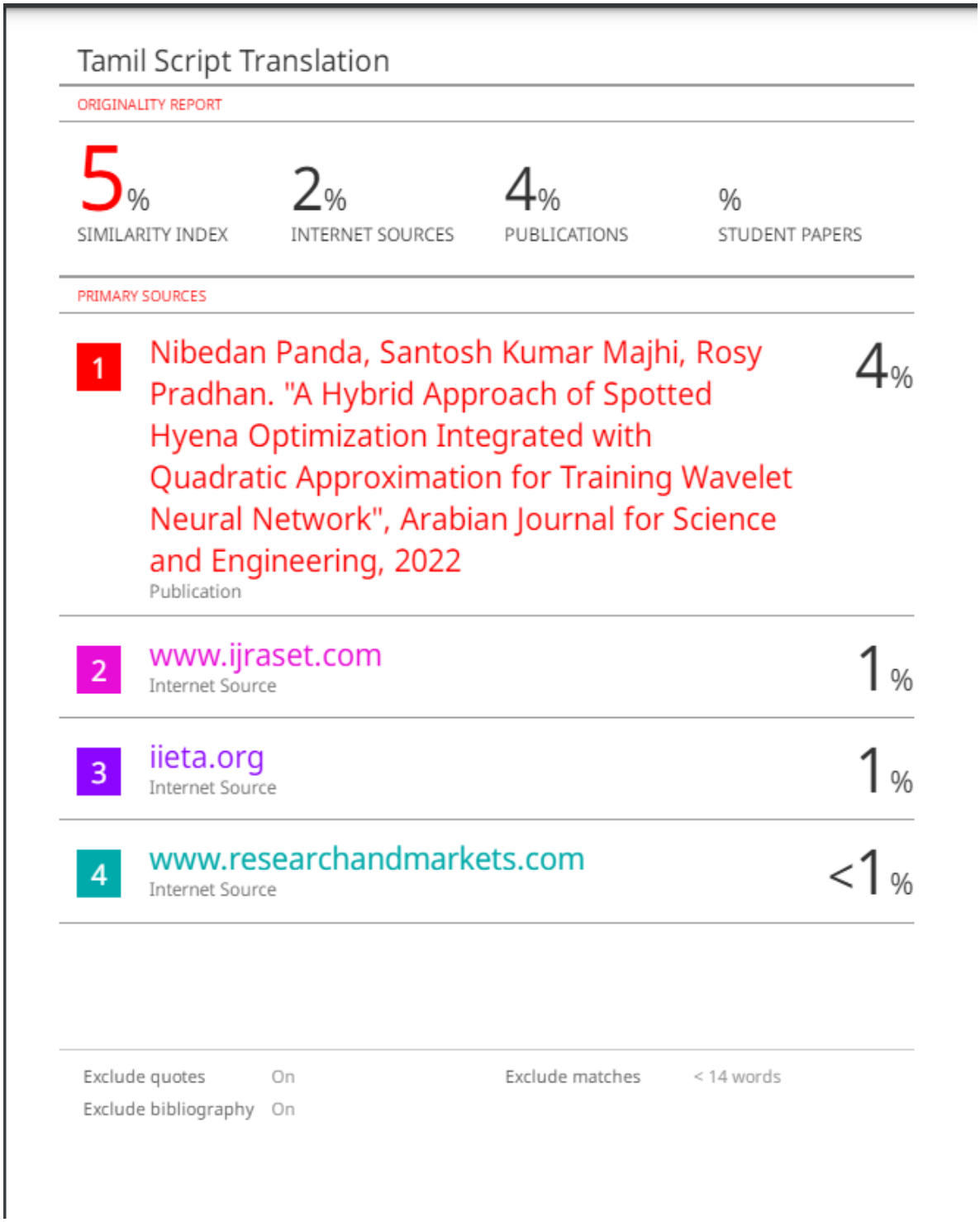
[4] R.Giridharan, E.K. Vellingiriraj and P. Balasubramanie-"Identification of Tamil Ancient Characters and Information retrieval from temple epigraphy using image zoning" (ICRTIT 2016).

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

[6] 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.Image recognition based billing system for fruit shop using raspberry PI A Vennila et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1055 012030

[7] 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.

Plagiarism Report



Tamil Script Translation

ORIGINALITY REPORT

5%	2%	4%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Nibedan Panda, Santosh Kumar Majhi, Rosy Pradhan. "A Hybrid Approach of Spotted Hyena Optimization Integrated with Quadratic Approximation for Training Wavelet Neural Network", Arabian Journal for Science and Engineering, 2022 Publication	4%
2	www.ijraset.com Internet Source	1%
3	iieta.org Internet Source	1%
4	www.researchandmarkets.com Internet Source	<1%

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		