OFFLINE HANDWRITTEN CHARACTER RECOGNITION OF ANCIENT MALAYALAM SCRIPT VATTEZHUTHU USING HYBRID VISION TRANSFORMER- SWIN MODEL

A Thesis Submitted to the University of Calicut

in partial fulfilment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

Under the Faculty of Science

By

AYYOOB MP

Under the guidance of

Dr. P Muhamed Ilyas

Principal (Rtd)

Sullamussalam Science College, Areekode

P.G & RESEARCH DEPARTMENT OF COMPUTER SCIENCE

Sullamussalam Science College, Areekode - 673639

(Affiliated to the University of Calicut)

Malappuram Dist., Kerala, India

November 2025

DECLARATION

I hereby declare that the work presented in the thesis entitled **OFFLINE HANDWRITTEN CHARACTER RECOGNITION OF ANCIENT MALAYALAM SCRIPT VATTEZHUTHU USING HYBRID VISION TRANSFORMER- SWIN MODEL** is based on the original work done by me under the guidance of **Dr. P Muhamed Ilyas** and has not been included in any other thesis submitted previously for the award of any degree. The contents of the thesis are undergone plagiarism check using Ithenticate software at C.H.M.K. Library, University of Calicut, and the similarity index found within the permissible limit. I also declare that the thesis is free from AI generated contents.

Signature

Name of the scholar AYYOOB MP

Signature of the Supervising teacher

Name Dr. MUHAMED ILYAS. P

Place: Areekode

Date: 12.11.2025

SULLAMUSSALAM
SCIENCE COLLEGE
AREACODE

Phone: 0483-2850700
Website: www.sscollege.ac.in
Email: mail@sscollege.ac.in

Aided College | Affiliated to the University of Calicut | Re-accredited by NAAC at A Grade

Ref: Date:

CERTIFICATE

This is to certify that the thesis entitled **OFFLINE HANDWRITTEN CHARACTER RECOGNITION OF ANCIENT MALAYALAM SCRIPT VATTEZHUTHU USING HYBRID VISION TRANSFORMER- SWIN MODEL** submitted by **Mr. AYYOOB MP**, PG and Research Department of Computer Science, Sullamussalam Science College Areekode, Malappuram for the award of the degree of Doctor of Philosophy, is a record of bonafide work carried out by him under my supervision, as per the Calicut University code of academic and research ethics. The contents of this thesis have not been submitted either in part or in full, for the award of any other degree or diploma in this Institute or any other Institute or University.

The thesis is revised as per the modifications and recommendations reported by the adjudicators. The soft copy attached is the same as that of the revised copy. The thesis is submitted as such to the University of Calicut with reference to the letter number No. 154486/RESEARCH-C-ASST-1/2025/Admn Dated 03.11.2025.

Dr. Muhamed Ilyas P.

Research Guide
Sullamussalam Science College,
Areekode

Place: Areekode

Date: 12/11/2025

SULLAMUSSALAM SCIENCE COLLEGE
UGRAPURAM P.O
MALAPPURAM DT.
PIN 673639
AREEKODE

UNIVERSITY OF CALICUT

CERTIFICATE ON PLAGIARISM CHECK

1.	Name of the Research Scholar	Ayyoob.MP	
2.	Title of thesis / dissertation	OFFLINE HANDWRITTEN CHARACTER RECOGNITION OF ANCIENT MALAYALAM SCRIPT VATTEZHUTHU USING HYBRID VISION TRANSFORMER- SWIN MODEL	
3.	Name of the Supervisor	Dr. Muhamed Illyas P	
4.	Department/Institution	Department of Computer Science, Sullamussalam Science College, Areekode	
5.		Non Core	Core
		Introduction/ Theoretical overview/Review of literature/ Materials & Methods/ Methodology	**Analysis/Result/Discussion/ Summary/Conclusion/ Recommendations**
	Similar content (%) identified	3	0
	Acceptable maximum limit (%)	10	10
6.	Software used	Ithenticate	
7.	Date of verification	06.06.25	

**Report on plagiarism check, specifying included/excluded items with % of similarity to beattached.*

Checked by (with name , designation & signature) Dr. Nasirudheen. T
Assistant Librarian
University of Calicut, Kerala.

Name and signature of the Researcher: Ayyoob.MP

Name and signature of the Supervisor: Dr. Muhamed Illyas P

The Doctoral Committee* has verified the report on plagiarism check with the contents of the thesis, as summarized above and appropriate measures have been taken to ensure originality of the Research accomplished herein.

Musthafa Farook.P
PRINCIPAL
SULLAMUSSALAM SCIENCE COLLEGE
AREEKODE,UGRAPURAM (PO)
MALAPPURAM(Dt),PIN-673639

Name & Signature of the HoD/HoI (Chairperson of the Doctoral Committee):

**In case of languages like Malayalam, Tamil etc..on which no software is available for plagiarismcheck, a manual check shall be made by the Doctoral Committee, for which an additional certificate has to be attached.*

ACKNOWLEDGMENTS

I am deeply grateful to all those who have contributed to the realization of this research endeavour. This journey has been both challenging and rewarding, and it is with immense gratitude that I acknowledge the support, guidance, and encouragement I have received throughout this process.

First and foremost, I wish to express my profound gratitude to my advisor, **Dr. Muhamed Ilyas P**, Associate professor of computer science Sullamussalam Science College Areekode. His unwavering support, insightful guidance, and scholarly expertise have been instrumental in shaping the direction of this research. His mentorship extended beyond academic instruction, offering invaluable encouragement and constructive feedback that not only refined my thesis but also enriched my intellectual growth. His patience, kindness, and positivity have been a source of inspiration, making this research journey a deeply rewarding experience.

I am equally indebted to **Prof. (Dr.) Mohamed Basheer K.P**, Research guide in the P.G. and Research Department of Computer Science at Sullamussalam Science College, for his unwavering support and encouragement. His leadership and dedication to academic excellence have inspired me to strive for the highest standards in my research pursuits. His mentorship has been pivotal in overcoming challenges and achieving my academic goals.

I extend my heartfelt gratitude to **Prof. N.V. Abdul Rahman**, the esteemed manager of Sullamussalam Science College, and the entire management team for their visionary leadership and support. Their commitment to providing the necessary resources and fostering a conducive academic environment greatly facilitated the successful completion of this research.

I am profoundly thankful to the esteemed members of my research advisory committee—**Dr. Lajish V.L.**, Associate Professor and Head, Department of Computer Science, University of Calicut; **Dr. Vasudevan**, Department of Library Science, University of Calicut; **Dr. Binu P. Chacko**, Associate Professor, Prajyothi Nikethan College Puthukkad; and **Dr. Shameem Kappan**, Head, P.G. and Research Department of Computer Science, Sullamussalam Science College—for their invaluable insights and constructive critiques. Their expertise and thoughtful guidance have greatly enriched my academic and intellectual development.

I would also like to express my gratitude to **Dr. S. Rajendu**, whose expertise and support were instrumental in navigating the complexities of this research. His positive influence and guidance have significantly contributed to the success of this work.

To my beloved parents and family, I owe my deepest gratitude. Their unwavering support, boundless love, and encouragement have been the foundation of my academic and personal achievements. I am forever grateful for their sacrifices and belief in my potential.

I extend my sincere appreciation to my colleagues, fellow scholars, and friends for their continuous encouragement and support throughout my academic journey. Their camaraderie and shared experiences have been a source of strength and motivation, making this journey a memorable and fulfilling one.

Finally, with humility and reverence, I thank the Almighty for His boundless blessings, guidance, and opportunities, without which this endeavour would not have been possible.

Ayyoob MP

Dedicated

To those who light the path as I seek Knowledge and Wisdom.

OFFLINE HANDWRITTEN CHARACTER RECOGNITION OF ANCIENT MALAYALAM SCRIPT VATTEZHUTHU USING HYBRID VISION TRANSFORMER- SWIN MODEL

ABSTRACT

Ancient scripts provide primary sources for understanding historical events, societies, and cultures that may not have been recorded elsewhere. They help establish timelines and sequences of events, offering a clearer picture of historical developments.

Among the numerous ancient scripts, Vattezhuthu stands out as one of the earliest in India, from the 8th to 15th centuries. This script contains information and knowledge that spans various fields, including history, culture, literature, law, science, mathematics, and medicine. However, its time-induced degradation, stylistic variability, and the scarcity of digitized samples pose significant challenges to preservation and study. This research addresses these obstacles by developing an automated framework for recognizing and digitizing Vattezhuthu script, integrating advanced image processing, innovative data augmentation, and a hybrid deep learning architecture. The study's primary objective is to enhance recognition accuracy while mitigating the limitations of degraded historical artifacts and insufficient datasets.

A comprehensive dataset was curated from stone inscriptions, copper plates, and palm leaf manuscripts sourced from repositories such as the Hill Palace Archaeological Museum, Tripunithura Palace, the State Archives Department and the University of Calicut. To address image degradation, a multi-stage preprocessing pipeline was implemented, including grayscale conversion, super-resolution techniques for detail enhancement, and noise reduction using median filtering and Gaussian smoothing. An adaptive binarization method was proposed, outperforming traditional algorithms (Otsu, Niblack, Sauvola) with high accuracy ensuring robust feature extraction from low-contrast, degraded manuscripts. The framework's efficacy was validated using metrics such as Peak Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE), and Structural Similarity Index Measure (SSIM). A novel stroke-based data augmentation technique was introduced to simulate natural handwriting variations, increasing dataset diversity and improving model generalizability.

For classification, the Hybrid Vision Transformer-Swin (HybridViTSwin) model was developed, combining the global self-attention mechanisms of Vision Transformers (ViTs) with the localized hierarchical attention of Swin Transformers. This architecture effectively captures both broad contextual patterns and fine-grained structural details of Vattezhuthu glyphs. Experimental results demonstrate the model's superiority, achieving 100% accuracy compared to standalone ViT (94.25%) and Swin (95.78%), confirming its robustness in handling stylistic and degradational complexities.

Dr. MUHAMED ILYAS P. AYYOOB MP

This work contributes theoretically through its hybrid attention mechanism and practically by releasing the publicly accessible Vattezhuthu dataset. Its implications extend beyond academia, offering museums and cultural institutions a scalable tool for digitizing endangered manuscripts. The framework's adaptability to other ancient scripts, such as Brahmi or Grantha, underscores its broader relevance. By bridging technological innovation with cultural preservation, this research not only safeguards a critical aspect of South Indian heritage but also establishes a replicable methodology for global historical script analysis.

Keywords

Vattezhuthu Recognition, Ancient Script Classification, Noise Reduction, Binarization Technique, Stroke-Based Data Augmentation, Vision Transformer and Swin Transformer.

AYYOOB M P

Dr. MUHAMED ILYAS P.

സംഗ്രഹം

മറ്റെവിടെയും രേഖപ്പെടുത്തിയിട്ടില്ലാത്ത ചരിത്രസംഭവങ്ങൾ, സമൂഹങ്ങൾ, സംസ്കാരങ്ങൾ എന്നിവ മനസ്സിലാക്കുന്നതിനുള്ള പ്രാഥമിക സ്രോതസ്സുകൾ പുരാതന ലിപികളാണ്. അവ സംഭവങ്ങളെ സമയാധിഷ്ഠിതമായി ക്രമപ്പെടുത്തി ചരിത്ര വികാസങ്ങളുടെ വ്യക്തമായ ചിത്രം പകർന്നു നൽകാൻ സഹായിക്കുന്നു.

നിരവധി പുരാതന ലിപികളിൽ, 8 മുതൽ 15 വരെ നൂറ്റാണ്ടുകളിൽ ഇന്ത്യയിലെ ഏറ്റവും പഴക്കമുള്ള ലിപികളിൽ ഒന്നായി വട്ടെഴുത്ത് വേറിട്ടുനിൽക്കുന്നു.

ചരിത്രം, സംസ്കാരം, സാഹിത്യം, നിയമം, ശാസ്ത്രം, ഗണിതം, വൈദ്യം തുടങ്ങി വിവിധ മേഖലകളിലായി വ്യാപിച്ചുകിടക്കുന്ന വിവരങ്ങളും അറിവും ഈ ലിപിയിൽ അടങ്ങിയിരിക്കുന്നു. എന്നിരുന്നാലും, അതിന്റെ കാലാനുസൃതമായ അപചയം, ശൈലീപരമായ വ്യതിയാനം, ഡിജിറ്റൈസ് ചെയ്ത സാമ്പിളുകളുടെ ദൗർലഭ്യം എന്നിവ സംരക്ഷണത്തിനും പഠനത്തിനും കാര്യമായ വെല്ലുവിളികൾ ഉയർത്തുന്നു. വട്ടെഴുത്ത് സ്ക്രിപ്റ്റ് തിരിച്ചറിയുന്നതിനും ഡിജിറ്റൈസ് ചെയ്യുന്നതിനുമുള്ള ഒരു ഓട്ടോമേറ്റഡ് ചട്ടക്കൂട് വികസിപ്പിച്ചുകൊണ്ട്, നൂതന ഇമേജ് പ്രോസസ്സിംഗ്, നൂതന ഡാറ്റ ഓഗ്മെന്റേഷൻ, ഒരു ഹൈബ്രിഡ് ഡീപ് ലേണിംഗ് ആർക്കിടെക്ചർ എന്നിവ സംയോജിപ്പിച്ചുകൊണ്ട് ഈ ഗവേഷണം ഇത്തരം തടസ്സങ്ങൾ പരിഹരിക്കുന്നു. തകർന്ന ചരിത്ര പുരാവസ്തുക്കളുടെയും, അപര്യാപ്തമായ ഡാറ്റാസെറ്റുകളുടെയും പരിമിതികൾ ലഘൂകരിക്കുന്നതിനൊപ്പം തിരിച്ചറിയൽ കൃത്യത വർദ്ധിപ്പിക്കുക എന്നതാണ് പഠനത്തിന്റെ പ്രാഥമിക ലക്ഷ്യം.

ഹിൽ പാലസ് പുരാവസ്തു മ്യൂസിയം, സംസ്ഥാന ആർക്കൈവ്സ് വകുപ്പ്, കാലിക്കറ്റ് സർവകലാശാലയിലെ തുഞ്ചൻ കയ്യെഴുത്തുപ്രതി ശേഖരം, തൃപ്പൂണിത്തുറ കൊട്ടാരം എന്നിവിടങ്ങളിൽ നിന്ന് ശേഖരിച്ച ശിലാ ലിഖിതങ്ങൾ, ചെമ്പ് ഫലകങ്ങൾ, താളിയോല കയ്യെഴുത്തുപ്രതികൾ എന്നിവയിൽ നിന്ന് സമഗ്രമായ ഒരു ഡാറ്റാസെറ്റ് നിർമ്മിച്ചെടുത്തു. ഇമേജ് ഡീഗ്രേഡേഷൻ പരിഹരിക്കുന്നതിനായി, ഗ്രേസ്കെയിൽ കൺവേർഷൻ, വിശദാംശങ്ങൾ മെച്ചപ്പെടുത്തുന്നതിനുള്ള സൂപ്പർ-റെസല്യൂഷൻ ടെക്നിക്കുകൾ, മീഡിയൻ ഫിൽട്ടറിംഗ്, ഗൗസിയൻ സ്മൂത്തിംഗ് എന്നിവ ഉപയോഗിച്ചുള്ള നോയ്സ് റിഡക്ഷൻ എന്നിവ ഉൾപ്പെടെ ഒരു മൾട്ടി-സ്റ്റേജ് പ്രീപ്രോസസിംഗ് പൈപ്പ്ലൈൻ നടപ്പിലാക്കി. പരമ്പരാഗത അൽഗോരിതങ്ങളെ (ഒറ്റ്സു, നിബ്ലാക്ക്, സൗവോള) മറികടന്ന് ഉയർന്ന കൃത്യതയോടെ കുറഞ്ഞ ദൃശ്യതീവ്രതയുള്ള, തരംതാഴ്ത്തപ്പെട്ട കയ്യെഴുത്തുപ്രതികളിൽ നിന്ന് ശക്തമായ സവിശേഷതകൾ വേർതിരിച്ചെടുക്കൽ ഉറപ്പാക്കുന്ന ഒരു അഡാപ്റ്റീവ് ബൈനറൈസേഷൻ രീതി നടപ്പിലാക്കി. പീക്ക് സിഗ്നൽ-ടു-നോയ്സ് റേഷ്യോ (പി.എസ്.എൻ.ആർ), മീൻ സ്ക്വയർ എറർ (എം.എസ്.ഇ), സ്ട്രക്ചറൽ സിമിലാരിറ്റി ഇൻഡക്സ് മെഷർ (എസ്.എസ്.ഐ.എം) തുടങ്ങിയ മെട്രിക്സുകൾ ഉപയോഗിച്ചാണ് ഫ്രെയിംവർക്കിന്റെ ഫലപ്രാപ്തി കണക്കാക്കിയത്. സ്വാഭാവിക കൈയക്ഷര വ്യതിയാനങ്ങൾ അനുകരിക്കുന്നതിനും, ഡാറ്റാസെറ്റ് വൈവിധ്യം വർദ്ധിപ്പിക്കുന്നതിനും, മോഡൽ സാമാന്യവൽക്കരണം മെച്ചപ്പെടുത്തുന്നതിനുമായി ഒരു നൂതനമായ സ്ട്രോക്ക് അധിഷ്ഠിത ഡാറ്റ ഓഗ്മെന്റേഷൻ സാങ്കേതികത അവതരിപ്പിച്ചു.

Dr. MUHAMED ILYAS P

വിഷൻ ട്രാൻസ്ഫോർമറുകളുടെ ഗ്ലോബൽ ഫീച്ചേഴ്സും സ്വിൻ ട്രാൻസ്ഫോർമറുകളുടെ ലോക്കൽ ഫീച്ചേർസും സമന്വയിപ്പിക്കുന്നതു കാരണം പ്രതീകങ്ങളെ തരംതിരിക്കുന്നതിന് ഏറ്റവും അനുയോജ്യമായ മോഡലായ വിഷൻ ട്രാൻസ്ഫോർമർ സ്വിൻ ഉപയോഗിച്ചാണ് വർഗ്ഗീകരണം നടത്തിയത്. പരീക്ഷണ ഫലങ്ങൾ മോഡലിന്റെ മികവ് തെളിയിക്കുന്നു, സ്റ്റാൻഡ്-എലോൺ ViT (94.25%), Swin (95.78%) എന്നിവയുമായി താരതമ്യപ്പെടുത്തുമ്പോൾ, 100% കൃത്യത കൈവരിക്കുന്നു, സ്റ്റൈലിസ്റ്റിക്, ഡീഗ്രേഡേഷണൽ സങ്കീർണ്ണതകൾ കൈകാര്യം ചെയ്യുന്നതിൽ അതിന്റെ കരുത്ത് സ്ഥിരീകരിക്കുന്നു.

സൈദ്ധാന്തികമായി ഈ ഗവേഷണം ഹൈബ്രിഡ് സംവിധാനത്തിലൂടെ പൊതുജനങ്ങൾക്ക് പ്രാപ്യമായ വട്ടെഴുത്ത് ഡാറ്റാസെറ്റ് സംഭാവന ചെയ്യുന്നു. നശിച്ചുകൊണ്ടിരിക്കുന്ന കയ്യെഴുത്തു പ്രതികൾ ഡിജിറ്റൈസ് ചെയ്യുന്നതിനുള്ള ഒരു വിപുലീകരിക്കാവുന്ന ഉപകരണം മ്യൂസിയങ്ങൾക്കും സാംസ്കാരിക സ്ഥാപനങ്ങൾക്കും വാഗ്ദാനം ചെയ്തുകൊണ്ട്, അക്കാദമിക് മേഖലയ്ക്കും അപ്പുറത്തേക്ക് അതിന്റെ സ്വാധീനം വ്യാപിക്കുന്നു. ബ്രാഹ്മി, ഗ്രന്ഥ തുടങ്ങിയ പുരാതന ലിപികളുമായി ഈ ചട്ടക്കൂടിന് ഇണങ്ങാനുള്ള കഴിവ് അതിന്റെ വിശാലമായ പ്രസക്തിയെ അടിവരയിടുന്നു. സാങ്കേതിക നവീകരണത്തെ സാംസ്കാരിക സംരക്ഷണവുമായി ബന്ധിപ്പിക്കുന്നതിലൂടെ, ഈ ഗവേഷണം ദക്ഷിണേന്ത്യൻ പൈതൃകത്തിന്റെ ഒരു നിർണായക വശം സംരക്ഷിക്കുക മാത്രമല്ല, ആഗോള ചരിത്ര ലിപി വിശകലനത്തിനായി ആശ്രയിക്കാവുന്ന ഒരു രീതിശാസ്ത്രം സ്ഥാപിക്കുകയും ചെയ്യുന്നു.

<u>കീവേഡുകൾ</u>

വട്ടെഴുത്ത് തിരിച്ചറിയൽ, പുരാതന ലിപി വർഗ്ഗീകരണം, നോയ്സ് കുറയ്ക്കൽ, ബൈനറൈസേഷൻ ടെക്നിക്, സ്ട്രോക്ക്-അധിഷ്ഠിത ഡാറ്റ ഓഗ്മെന്റേഷൻ, വിഷൻ ട്രാൻസ്ഫോർമറും സ്വിൻ ട്രാൻസ്ഫോർമറും.

Dr. MUHAMED ILAYAS P

SULLAMUSSALAM SCIENCE COLLEGE
UGRAPURAM P.O
MALAPPURAM DT.
PIN-673639
AREEKODE

Table of Contents

1. INTRODUCTION ---1

1.1 BACKGROUND AND CONTEXT ---1

1.2 EVOLUTION AND CHARACTERISTICS OF THE VATTEZHUTHU SCRIPT---2

1.3 MOTIVATION---5

1.4 PROBLEM SPECIFICATION --6

1.5 AIM AND OBJECTIVE---8

1.6 OUTLINE OF THE THESIS--9

1.7 CONCLUSION--- 10

2. LITERATURE REVIEW -- 11

2.1. PRE-PROCESSING BASED RESEARCH ------------------------------------ 11

2.1.1 INTRODUCTION -- 11

2.1.2 REVIEWS ON NOISE REMOVAL TECHNIQUES -------------------- 11

2.1.3 BINARIZATION TECHNIQUES -- 12

2.1.3.1 INTRODUCTION --- 12

2.1.3.2 REVIEWS ON BINARIZATION TECHNIQUES ------------------ 12

2.2 SEGMENTATION TECHNIQUES --- 13

2.2.1 INTRODUCTION -- 13

2.2.2 REVIEWS ON SEGMENTATION TECHNIQUES --------------------- 13

2.3 DATA AUGMENTATION TECHNIQUES---------------------------------- 20

2.3.1 INTRODUCTION -- 20

2.3.2 REVIEWS ON DATA AUGMENTATION TECHNIQUES----------- 21

2.4 CLASSIFICATION TECHNIQUES --- 23

2.4.1 INTRODUCTION -- 23

2.4.2 CLASSIFICATION BASED REVIEW------------------------------------ 24

2.4.3 CONCLUSION-- 66

2.5 VISION TRANSFORMER CLASSIFIER -- 67
2.5.1 INTRODUCTION -- 67
2.5.2 VISION TRANSFORMER BASED REVIEW ---------------------------- 68
2.5.3 CONCLUSION-- 78
3. RESEARCH METHODOLOGY --- 79
3.1 INTRODUCTION --- 79
3.2 DATASET CONSTRUCTION -- 79
3.3 GRAY SCALE CONVERSION -- 79
3.4 SUPER RESOLUTION (SR) -- 79
3.5 NOISE REDUCTION TECHNIQUES --- 80
3.6 BINARIZATION -- 80
3.7 SEGMENTATION --- 80
3.8 RESIZING NORMALIZATION -- 80
3.9 DATA AUGMENTATION-- 80
3.10 HYBRID VISION TRANSFORMER-SWIN MODEL CLASSIFICATION--- 81
4. PHASES OF DATA COLLECTION -- 82
4.1 INTRODUCTION --- 82
4.2 STRATEGIES FOR DATASET CONSTRUCTION -------------------------- 82
4.3 GRAY SCALE CONVERSION -- 83
4.4 SUPER RESOLUTION (SR) -- 84
5 NOISE REDUCTION TECHNIQUES -- 85
5.1 INTRODUCTION --- 85
5.2 REMOVAL OF SPECKLE NOISE -- 85
5.3 REMOVAL OF SALT AND PEPPER NOISE -------------------------------- 85
5.4 REMOVAL OF GAUSSIAN NOISE--- 86
6 BINARIZATION --- 87
6.1 INTRODUCTION --- 87
6.2 LOCAL OTSU --- 87

6.3 NIBLACK-- 88
6.4 SAUVOLA -- 88
6.5 PROPOSED METHOD- GAUSSIAN ADAPTIVE THRESHOLDING WITH EDGE-ENHANCED DILATION (GATED) ------------------------------ 89
6.6 EXPERIMENTAL RESULT AND PERFORMANCE EVALUATION--- 93
6.7 CONCLUSION-- 95
7 SEGMENTATION AND NORMALIZATION ------------------------------------ 97
7.1 INTRODUCTION --- 97
7.2 PROPOSED METHOD--- 97
7.3 CONCLUSION-- 98
7.4 RESIZING NORMALIZATION -- 99
8. DATA AUGMENTATION ---100
8.1 INTRODUCTION ---100
8.2 PROPOSED METHOD- STROKE BASED AUGMENTATION (SBA) 101
8.2.1 LOCALIZED THICKENING AND THINNING OF STROKES ---101
8.2.2 THICKENING AND THINNING BASED ON STROKE DIRECTION ---105
8.2.3 COMPARATIVE ANALYSIS --112
8.2.4 CONCLUSION---113
9 HYBRID VISION TRANSFORMER-SWIN MODEL FOR ROBUST AND ACCURATE CLASSIFICATION ---117
9.1 INTRODUCTION ---117
9.2 VIT-TINY (VISION TRANSFORMER) [VIT_TINY_PATCH16_224] --119
9.3 SWIN-BASE TRANSFORMER [SWIN_BASE_PATCH4_WINDOW7_224] ---------------------------------------121
9.4 PROPOSED MODEL (HYBRIDVITSWIN) ---------------------------------124
9.5 EXPERIMENT RESULT AND PERFORMANCE EVALUATION------127
9.6. CONCLUSION ---129
10. RESULTS AND ANALYSIS ---131
11. CONCLUSION--132

12. CONTRIBUTIONS --133
13. IMPLICATIONS --134
14. LIMITATIONS AND FUTURE WORK --135
References---136
List of Publications Out of Thesis Work ---153
Conference Paper Presentations--154

List of Figures

Figure 1 Geographical distribution of writing systems in the Indian subcontinent across different historical periods ------------------------------------4
Figure 2 Basic Character Set of Vattezhuthu Script ---------------------------------5
Figure 3 Permission letter from Kerala State Archive Department to consult records--7
Figure 4 Permission letter from Hill Palace archaeological museum and palace Tripunithura Kerala to consult records. --8
Figure 5 Permission letter from university of Calicut Tunjan manuscript repository to consult records ---8
Figure 6 Original images and its corresponding Gray scaled images--------- 84
Figure 7 Binarized stone inceptions of the above methods ---------------------- 91
Figure 8 Binarized copper plate of the above methods --------------------------- 92
Figure 9 Binarized palm leaf of the above methods ------------------------------- 92
Figure 10 PSNR, MSE and SSIM values of the copper plate, stone ------------ 95
Figure 11 Results of characters segmented from the sample input image --- 98
Figure 12 a) Binarized input image 'pa'. b) The thinned regions were ------102
Figure 13 Shows the thickening of the stroke at different regions of---------103
Figure 14 Characters in each set ---107
Figure 15 Thinning and thickening of characters in Set-1 -----------------------107
Figure 16 Thinning and thickening of characters in Set-2 -----------------------108
Figure 17 Thinning and thickening of characters in Set-3 -----------------------109
Figure 18 Thinning and thickening of characters in Set-4 -----------------------109
Figure 19 Thinning and thickening of characters in Set-5 -----------------------110
Figure 20 vit_tiny_patch16_224 model ---119
Figure 21 confusion matrix of vit_tiny_patch16_224 model --------------------120
Figure 22 Swin Transformer Model ---122
Figure 23 Confusion matrix of swin_base_patch4_window7_224. -----------123
Figure 24 Architecture of proposed method ---------------------------------------125
Figure 25 Confusion matrix of proposed method --------------------------------126

List of Tables

Table 1 Character recognition in research 50
Table 2 Specifications of the image captured device 83
Table 3 PSNR, MSE and SSIM values of the copper plate, stone and palm leaf images binarized by various techniques 94
Table 4 Direction of stroke and corresponding characters set 105
Table 5 Comparison of traditional data augmentation techniques 113
Table 6 Comparison of traditional data augmentation and proposed data augmentation techniques...114
Table 7 Number of characters in each class after the augmentation 116
Table 8 vit_tiny_patch16_224 performance..120
Table 9 swin_base_patch4_window7_224 performance 123
Table 10 Comparison of the proposed model with the CNN-based architectures..127
Table 11 Comparison of the proposed model with hybrid transformer models 127
Table 12 Performance Comparison metric 128

List of Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural network
CoT	Context-Transformer
GAN	Generative Adversarial Network
MSE	Mean Square Error
NST	Neural Style Transfer
OCR	Optical Character Recognition
OHR	Offline Handwriting Recognition
PDA	Personal Digital Assistant
PDF	Probability Density Function
PSNR	Peak Signal-to-Noise Ratio
RGB	Red Green Blue
ROI	Region Of Interest
SR	Super Resolution
SSIM	Structural Similarity Index Measure
TCNet	Transformer-Convolution Network
VGG	Visual Geometry Group
ViTs	Vision Transformers
ViT-CNN	Vision Transformer with Convolutional Neural network

1. INTRODUCTION

1.1 BACKGROUND AND CONTEXT

The technology known as Optical Character Recognition (OCR) transforms printed or written documents into electronic or machine-readable text, making it possible to recognize handwritten text automatically. The automatic recognition of handwritten text falls into two main categories: offline and online handwriting recognition. Transferring fixed handwritten text from picture form into typed text is known as Offline Handwriting Recognition (OHR). This method views a handwritten piece as immutable, unlike interpretable data, which determines its meaning by relocating a pen over paper, like picking or dropping a pen. On the other hand, online handwriting recognition uses special devices like Personal Digital Assistants (PDA) and digitizers to capture motion data. The so-called dynamic data, also known as digital ink, significantly enhances recognition accuracy by providing a vivid account of the handwriting process.

For one, offline handwriting recognition is much harder because there is no instantaneous data concerning the pen's motion. Globalization is driving an increasing need to organize and search handwritten text in a format accessible from any physical location. Rapid technological advancements in image processing have created a significant demand for OCR systems, especially for handwritten documents.

Dealing with ancient manuscripts particularly highlights the offline OCR exigencies. These documents experience various forms of degradation over time as they get old, such as faded ink, smudging, thin or faint letters, blurring, and even uneven light [1]. Furthermore, the imaging device, whether a camera or a scanner, determines the recognition of these texts. Scanners typically provide higher-quality images, while cameras are more prone to blurring, resulting in reduced image quality and errors in recognition. Among the numerous ancient scripts, Vattezhuthu stands out as one of the earliest in India. Kerala used this script for administration and commerce from the 8th century to the 15th century, during a period when the area spoke Tamil. We should preserve the Vattezhuthu script because it holds significant historical and cultural value. Nevertheless, the joined letters found in Vattezhuthu and the gaps between some of them make the automation of writing recognition systems difficult. The situation is further worsened by the state of old manuscripts, which are marked by soaking ink, fading characters, and a lack of illumination. This condition is essential since overcoming these challenges will guarantee a faithful reading, reproduction, and digital preservation of the script.

This policy aims to improve Vattezhuthu offline recognition by making it more general, fixing its primary problem, and using up-to-date OCR technology. By resolving these issues, these efforts contribute to the preservation of a vast literary culture and aid in utilizing ancient ideas stored within the script.

1.2 EVOLUTION AND CHARACTERISTICS OF THE VATTEZHUTHU SCRIPT

Palaeography, defined as the discovery of ancient writing styles, is crucial for understanding the development of people and civilizations' language, culture, and society. Through the study of manuscripts, scrolls, and stone carvings, palaeography is able to highlight aspects of language change, language use, and cultural practices through communication that influenced the history of mankind. Its focal point is centred on the origin, evolution, and adoption of different scripts, paying attention to the civilizations that developed them.

One of the prominent scripts studied by South Indian palaeographers revolves around the Vattezhuthu, a widely used script in Kerala and Tamil Nadu. Vattezhuthu can be translated to round script which features cursive rounded forms of letters most likely derived from the southern Brahmi script. Vattezhuthu was in use from the 8^{th} to 15^{th} century AD and it was widely used to write in Tamil and in the early forms of Malayalam as used in administrative, literary and religious writings.

In the past, every written language was developed by language communities around the globe, and it was essential to the people's way of life. The diversity in the writing systems led to different scripts, whereby communities used the script to identify themselves and their languages. The Vatteluttu script, on the other hand, was one of the writing systems found on the Malabar Coast and Tirunelveli, as well as in the Madurai area, and was used significantly.

In South India, the Tekkan-Malayakm or Nana-Mona, otherwise known as Vatteluttu script, was prevalent. The reign of Tamil Pandya King Jatilavarman Parantaka I, around the late 8th century AD, produced the oldest surviving writings in this language. In those days, the Vatteluttu script had nearly reached its final form with clear, easy-to-read, and understandable letters.

Neglect, through the years, resulted in a decline in the quality standard of writing. This degeneration resulted in the increasing indistinctiveness of letters. Nevertheless, upon this degeneration taking place, the Vatteluttu script was still in use along the Western Coast until the close of the 18th century, after which it was gradually supplanted by the Malayalam script.

As for the questions on when Vatteluttu first appeared, they have been discussed by multiple scholars. Dr. Burnell theorized that the script may have been influenced by the Phoenician and European Aramaic letters. Nonetheless, he argued that the best approach to seeing the Asoka inscriptions and the Vatteluttu is to consider them as different writing systems. Dr. Burnell regarded Vatteluttu as being long antedating the Tamil writing system and maintained that primitive forms of Tamil work, especially those of Tolkappiyam and Tirukkural, were composed using this script. Some of his conclusions concerning the age of the content of Tamils' classical literature were, for instance, questioned by the Thamizh pundit M.Sundaram Pillai in his Thamizh IlakiyamAlaiyankal Paniyut Heathukal.

Epigraphy, which examines inscriptions, was still a developing area of research as Dr. Burnell was writing, and this meant that it was impossible to be certain as to the attributes and development of writers such as Vatteluttu. Progress in the development of ideas, particularly in the work of historians like Dr. Buhler in his Indian Palaeography, enabled this to be better understood, particularly the script. Dr. Buhler characterized Vatteluttu as evolving as a cursive hand that was relatively influenced by the Tamil people but independently developed [2].

The Vatteluttu script is essential to our understanding of the development of writing in southern India and how this development relates to other languages and cultures. Vatteluttu is a hybrid and a distinct Indian script and therefore occupies an important place within the history of writing within the area.

Figure 1 represents the evolution and geographical distribution of writing systems in the Indian subcontinent across different historical periods. The timeline depicted emphasizes the continuous evolution of writing systems in India, beginning with the undeciphered Indus script, followed by the development of Brahmi as a foundational script, and later regional adaptations like Vatteluttu and Tamil.

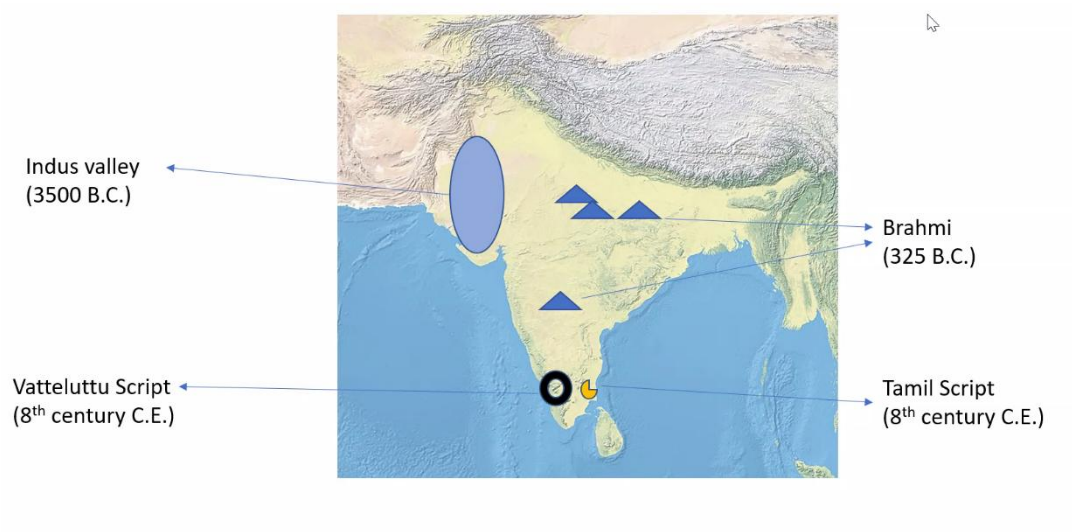


Figure 1 Geographical distribution of writing systems in the Indian subcontinent across different historical periods

The co-existence of scripts such as Vatteluttu and Tamil during the 8th century shows the dynamic nature of language and script evolution, shaped by cultural, administrative, and geographical influences.

The major writing systems of a language express the peculiarities of the countries and historical epochs they lived through. A script evolves into another, thus producing a new writing system, and at some times, there could be several scripts in use at the same time as well. Such phenomena are evident in the linguistic history of Kerala where even though existed in different times Vattezhuthu, Kolezhuthu, Grantha and Malayanma scripts were in use.

Scholars are, however, more or less agreed in their attribution of the Vattezhuthu's origins to the southern branch of the Brahmi script. It was widely used in regions of Tamil Nadu, the Pandya kingdom and the Malanad region in Kerala. In the border regions of Tamilnadu and Kerala, there are inscriptions in Vattezhuthu and Grantha revealing the influence of these writing systems on each other.

It was clear that Brahmins used Grantha to write Prakrit and Sanskrit from the observations of the Mampally inscriptions (Malayalam Era 2nd century-AD 10th century). This script was used with Tamil and Vattezhuthu in the Malanad area, particularly for Sanskrit words. One important example is the

Thiruvampadi inscription of the 12th century A.D. The first six lines are scripted in Grantha, and the balance is in Vattezhuthu.

By the 18th century, the use of Vattezhuthu, Kolezhuthu, and Malayanma became common in copper plate inscriptions, manuscripts on palm leaves, and sometimes on papers with only slight variations in areas and style. Although these scripts have originated at different times, they were used in conjunction and concurrently till the end of the nineteenth century, and this was a true picture of the linguistic scenario of Kerala. Vattezhuthu contains nine vowels and eighteen consonants as shown in figure 2.

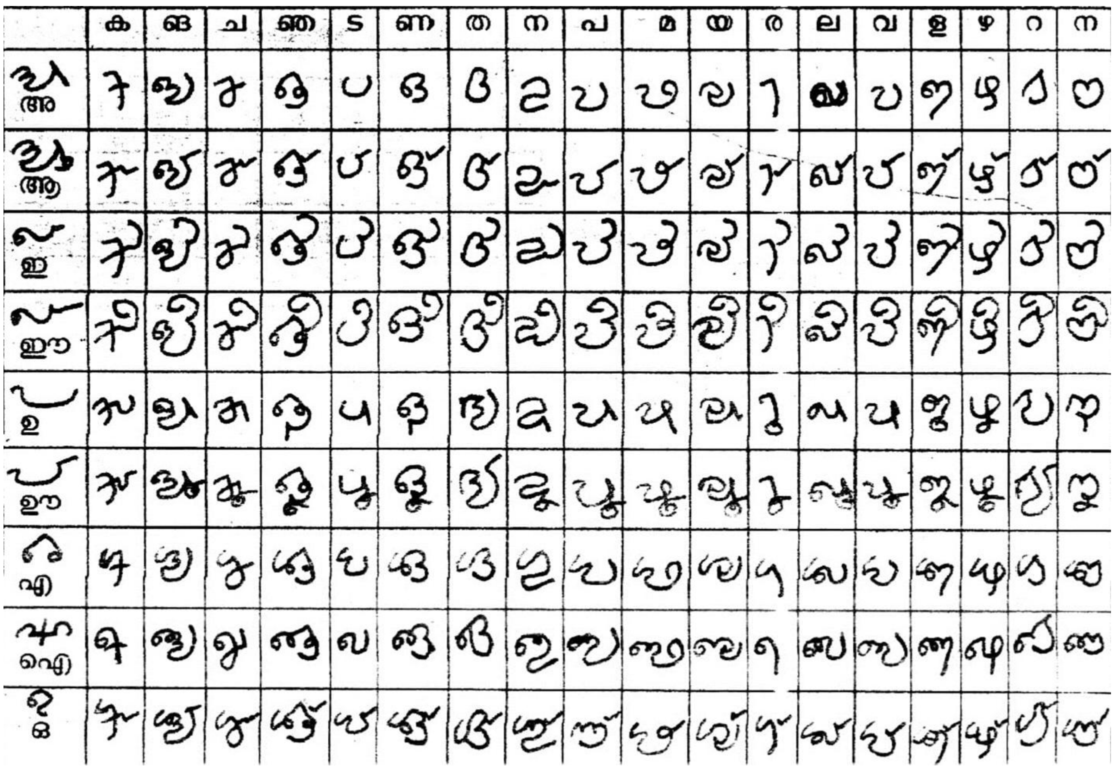


Figure 2 Basic Character Set of Vattezhuthu Script

1.3 MOTIVATION

Several motivations drive the investigation of the ancient Vattezhuthu script of Malayalam. Most importantly, scripts preserve and inform society's history and culture. Nevertheless, the Vattezhuthu script is quite poorly available; thus, the requirement for its conservation and research is dire.

Abundant Repository of History and Culture:

The Vattezhuthu script functions as a vital archival source for the culture and history of the region.

Limited Availability:

The evident limited use of the Vattezhuthu script poses difficulties for researchers and scholars in accessing and studying it.

Diverse Recording Materials:

Historically, the script was recorded on various materials such as stone, copper plates, and palm leaves, reflecting its adaptability and significance in ancient times.

Labor Intensive Conventional Methods:

Traditional methods of copying the manuscript are quite time-consuming and inefficient, which makes thorough study and recording extremely difficult.

Low Artifact Quality:

Low-quality imprints originating from traditional means also cause issues in the digitization and depository processes.

Challenges in Preservation:

The instability of the manuscript contributes to the difficulty of maintaining the remaining scripts in their original form.

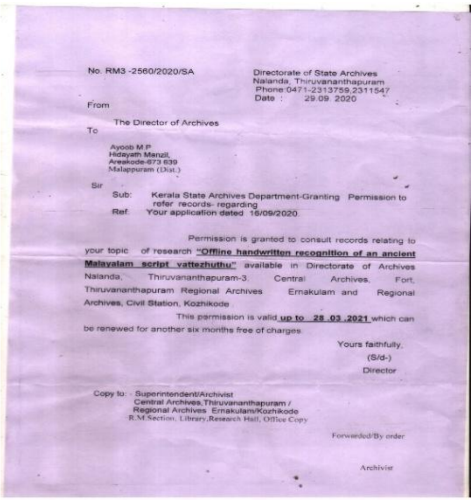
1.4 PROBLEM SPECIFICATION

Recognizing handwritten characters is inherently difficult because, in most writing systems, the characters have well-defined sizes and a defined position on the page. It is practically impossible. As the outlines of the problem get more intricate, so too do the structures of the language's characters and the styles in which they are written. The Vattezhuthu script is a good example of this because the lettering is ancient, the cases are tough, and the structures are often mixed up, which makes the classification process both time-consuming and complex.

In this regard, the case of Vattezhuthu is particularly difficult because datasets are abundant. Furthermore, the lack of publicly available datasets for this type of Vattezhuthu at the line level has significantly hindered the advancement of recognition research on the Vattezhuthu script.

This gap is filled by creating a corpus of Vattezhuthu characters. For this purpose, samples were collected from historical resources like copper plates and stone inscriptions acquired from the Hill Palace Archaeological Museum and the palace within the Tripunithura area of Kochi, Kerala, India (see figure 3). Furthermore, palm leaf manuscripts were fetched from the State Archives Department, Thiruvananthapuram, Kerala, India, and the University of Calicut Tunjan Manuscript Repository (see figure 4 and figure 5). Identifying handwritten characters poses significant challenges due to the rigid structural arrangements of characters in most scripts. Without the guidance of a curator, the accurate recognition of characters becomes nearly impossible. The complexity of the problem increases with the unique outlines and intricate structures of linguistic characters, as well as varying writing styles. This challenge is particularly evident in the case of the Vattezhuthu script, an ancient script characterized by rigid structures and confusing outlines, making its classification both a complicated and demanding task.

Additionally, the writing styles for the same character differ between individuals and may even vary for a single individual due to factors such as aging, writing attitude, and applied force. Handwriting clarity further impacts the accuracy of character recognition. Environmental factors exacerbate these issues, with poor storage conditions, paper decay, and ink seepage accelerating the deterioration of source materials.



No. RM3 -2560/2020/SA

Directorate of State Archives
Nalanda, Thiruvananthapuram
Phone:0471-2313759,2311547
Date : 29.09. 2020

From

The Director of Archives

To

Ayoob M.P
Hidayath Manzil,
Areakode-673 639
Malappuram (Dist.)

Sir

Sub: Kerala State Archives Department-Granting Permission to refer records- regarding

Ref: Your application dated 16/09/2020

Permission is granted to consult records relating to your topic of research **"Offline handwritten recognition of an ancient Malayalam script vattezhuthu"** available in Directorate of Archives Nalanda, Thiruvananthapuram-3, Central Archives, Fort, Thiruvananthapuram Regional Archives Ernakulam and Regional Archives, Civil Station, Kozhikode

This permission is valid **up to 28 .03 .2021** which can be renewed for another six months free of charges

Yours faithfully,
(S/d-)
Director

Copy to: Superintendent/Archivist
Central Archives,Thiruvananthapuram /
Regional Archives Ernakulam/Kozhikode
R.M.Section, Library,Research Hall, Office Copy

Forwarded/By order

Archivist

Figure 3 Permission letter from Kerala State Archive Department to consult records

File No.ARCDA/1540/2022-T5

Azadi Ka Amrit Mahotsav

പുരാവസ്തു വകുപ്പ്

സുന്ദരവിലാസം കൊട്ടാരം
തിരുവനന്തപുരം - 23
email : cru.archkerala.gov.in
തീയതി: 13/01/2023

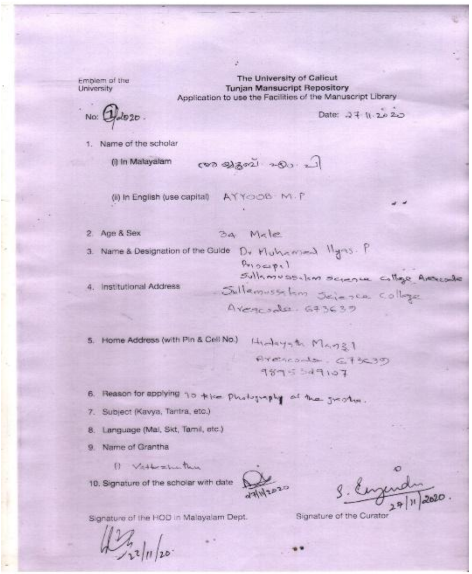
mpayyoobmp@gmail.com

2. വിവരശേഖരണത്തിനുള്ള തിയതി ചാർജ്ജ് ഓഫീസറുമായി ബന്ധപ്പെട്ട് മുൻകൂട്ടി നിശ്ചയിക്കേണ്ടതാണ്. (0484 – 2781113)

3. സംരക്ഷിത വസ്തുവിന് യാതൊരുവിധ കേടുപാടുകളും വരുത്താൻ പാടുള്ളതല്ല.

Signature Not Verified
Digitally signed by BEENA U K
Date: 2023.01.1? 12:32:32 IST
Reason: Approved

Figure 4 Permission letter from Hill Palace archaeological museum and palace Tripunithura Kerala to consult records.



Emblem of the University

The University of Calicut
Tunjan Mansucript Repository
Application to use the Facilities of the Manuscript Library

No: 1/2020. Date: 27.11.2020

1. Name of the scholar
   (i) In Malayalam
   (ii) In English (use capital) AYYOOB M.P
2. Age & Sex 34 Male
3. Name & Designation of the Guide Dr Muhammed Ilyas. P Principal Sullamussalam Science College Areacode
4. Institutional Address Sullamussalam Science College Areacode. 673639
5. Home Address (with Pin & Cell No.) Hidayath Manzil Areacode. 673639 9895549107
6. Reason for applying To take Photography of the grantha.
7. Subject (Kavya, Tantra, etc.)
8. Language (Mal, Skt, Tamil, etc.)
9. Name of Grantha
10. Signature of the scholar with date 27/11/2020

Signature of the HOD in Malayalam Dept. 27/11/20

Signature of the Curator 27/11/2020.

Figure 5 Permission letter from university of Calicut Tunjan manuscript repository to consult records

1.5 AIM AND OBJECTIVE

The goal of this study is to create an automated character transliteration system that can find and process relevant scripts that hold specific information. This will help researchers process their knowledge. This system aspires to become a pivotal tool for the collection of ancient data and enable the extraction of thematic information from the wisdom of ages.

The objectives of the research are:

Primary Objective:
Design and develop an efficient character recognition model for offline handwritten Vattezhuthu script, enhancing recognition accuracy using soft computing techniques and emerging learning methodologies.

Database Development:
Create a comprehensive database of the Vattezhuthu script by collecting samples from historical sources.

Noise Reduction:
In order to minimize the negative impacts of noise, various noise reduction techniques such as median filtering, despeckle filtering, and Gaussian smoothing are utilized.

Binarization Methodology:
Implement an adaptive binarization method named GATED (Gaussian Adaptive Thresholding with Edge-Enhanced Dilation) to effectively binarize thin and highly degraded ancient Malayalam script Vattezhuthu for improved processing quality.

Data Augmentation

Presents a novel stroke-based data augmentation technique called SBA (Stroke Based Augmentation) for enhancing the recognition accuracy of Vattezhuthu characters.

Classification Algorithm:
Develop a novel classification algorithm named HybridViTSwin, which integrates the strengths of both Swin-Base and vision transformer models to address the complexities of classifying ancient Vattezhuthu scripts efficiently.

1.6 OUTLINE OF THE THESIS

This thesis consists of 14 chapters. The reference of the study is given at the end of the thesis, followed by an appendix and a list of publications.

1.7 CONCLUSION

Handwriting systems are challenging to decipher, especially for the less frequently used older systems like Vattezhuthu. This is primarily because these scripts are complex in their structuring; people write the same script in different styles, and the available datasets are not abundant. The objective of this research was to bridge this void by compiling a complete set of the Vattezhuthu letters available on antique objects such as copper plates, stone epigraphs, and palm leaf manuscripts in the particular region of Kerala in India.

To understand the problems with Vattezhuthu character recognition, the study suggested a strong framework with noise reduction, hybrid binarization techniques, new data enhancement methods, and an advanced classifier called the HybridViTSwin classifier, which was made more accurate at recognition. The research makes substantial progress in recognizing ancient scripts by integrating various soft computing approaches with advanced machine learning techniques.

2. LITERATURE REVIEW

2.1. PRE-PROCESSING BASED RESEARCH

2.1.1 INTRODUCTION

Old handwritten documents provide a deeper contextual understanding of cultural history, but only if they are preserved well. The tangible and time-related decay, noise, and other deficiencies make interpreting old documents difficult. It is crucial to have digital preservation and interpretation so that these historical imprints retain their significance in the future.

Several algorithms aim to preserve historical text and improve readability without changing the context or meaning.

This literature review explains the evolution of methods and current practices in the analysis of historical documents. Their predominant aim has always been to safeguard cultural heritage and improve the quality of these records, thus making ancient texts more accessible to contemporary society. This literature review also looks at past research and shows how quickly and significantly denoising methods are improving the preservation of historical documents.

2.1.2 REVIEWS ON NOISE REMOVAL TECHNIQUES

Handwritten records made centuries ago are likely to contain different types of noise owing to their age [3], the ambient conditions they have been exposed to, and the limitations of the means of preserving historical records. In this regard, such noise might interfere with the text's intended meaning and may even alter content analysis. To counter this, many denoising algorithms have been proposed in the literature concerning the kind of historical documents embraced as an example. How to denoise and enhance palm-leaf manuscripts has been tackled in various publications [4,5,6,7,8], as well as treating copper plate engravings [9], and stone inscriptions [10].

These methods aim to make ancient texts easier to read and more coherent, improving knowledge and preservation. The ongoing evolution and improvement of these techniques for removing noise are important for the progress of historical document analysis and cultural heritage protection.

2.1.3 BINARIZATION TECHNIQUES

2.1.3.1 INTRODUCTION

A new field of study has emerged, which is concerned with the digitalization and archiving of historical texts, scripts, and inscriptions, as such artifacts often deteriorate and have irregular surfaces and noise, making them challenging to study. This technique is essential for reading and analysis, as it involves converting grayscale images to binary images through trends. So, the researchers have always been trying to explore and develop different techniques, including binarization and image enhancement based on the specific nature of the historical documents and inscriptions.

This literature review covers specific techniques to improve binarization and image processing. Some materials, such as stone inscriptions, copper plates, and palm leaf manuscripts, are often considered unique or special. When conventional methods of handling these materials are not up to scratch, fastidious approaches have developed for accomplishing such tasks. This review investigates recent developments, experimental findings, and essential adaptive binarization techniques to demonstrate how they assist in heritage conservation and indicate what further steps in the domain require.

2.1.3.2 REVIEWS ON BINARIZATION TECHNIQUES

A couple of researchers have explored more complex approaches to addressing these issues. Through the H-DIBCO 2012 dataset, Mustafa et al. [11] investigated seven different binarization techniques and noted that the gradient-based method was the most effective.

Correspondingly, Omar Boudraa et al. [12] investigated several thresholding techniques to identify the most effective ones. Rasmana et al. [13] developed an amalgamated strategy incorporating Otsu thresholding and Gray Level Cooccurrence Matrix (GLCM) features. Their resulting conclusion was impressive, with an F-measure of 95.4%, pseudo-F-measure of 94.44%, and PSNR of 11.15.

S. Das et al. [14] explored substituting noise in images of stone inscriptions by applying linear regression and Fast ICA (Independent Component Analysis) components analysis. They determined threshold values by using cumulative residual entropy.

For stone writings in Tamizhi script, Monisha Munivel and V.S. Felix Enigo [15] utilized a multi-level modified binarization technique. Some other authors, S. Raja Kumar et al. [16], proposed an effective method for obtaining Tamil sounds from stone inscriptions, and Bhuvaneswari et al. [17] devised new methods for imaging ancient stone inscriptions and enhancing images.

Also, many techniques were created that have the capability of enhancing images or even binarizing those images, thus making palm leaf manuscripts legible. There are works [18-23] which explain procedures that increase the accuracy of binarization, which is a nice way of solving the problems associated with uneven textures, degradation, and noise of palm leaf documents. The importance of adaptive and innovative approaches in safeguarding and analysing ancient texts is stressed by these advances, which set the stage for additional work in the field.

2.2 SEGMENTATION TECHNIQUES

2.2.1 INTRODUCTION

Segmentation approaches are very relevant for a variety of computer vision and pattern recognition tasks, especially character recognition and language processing. Such applications include the digitalization of historical documents, real-time vehicle plate recognition and intelligent transport systems. There is a myriad of segmentation novel techniques developed around various scripts, and multi-lingual handwritten texts.

This literature review is focused on the evaluation of segmentation techniques using classical approaches, as well as machine and deep learners. The discussion brings out the need for segmentation as a vital aspect required before performing other tasks such as classification, recognition, and repair. All analysed work peers into different application areas which hold different problems such as: structural complexity, overlapping characters, different writing scripts, and obscure or damaged text.

The review stresses the importance of a variety of datasets in model training and their resilience in segmentation model development.

2.2.2 REVIEWS ON SEGMENTATION TECHNIQUES

Kunnapat Thipparaphonkul et al. [28] proposed a new strategy for offline segmentation of handwritten Thai characters. They restored this method using deep learning techniques and transformed the convolutional encoder-decoder structure. They built a special type of segmentation named U-SegNet, which is a hybrid combination of features from U-Net and SegNet.

The authors were keen on the fact that a large and somewhat heterogeneous dataset was required to successfully train these deep learning models. To ensure that there is variation in writing styles and character combinations, 5614 handwritten character images were sourced from over 50 writers and combined into a single dataset. They also used synthetic validation sets in conjunction with the source set to reduce overfitting and bias. They achieved enhanced model outputs using this method, which led to an exponential increase in character recognition accuracy.

The research points out the important role of character classification in effective segmentation and recognition. The authors developed segmentation models by constructing four distinct character locations: the topmost, higher, base, and bottom to accommodate the structural complexity of Thai handwriting. Their results showed remarkable performance, as it was noted that the horizontal link segmentation model performed best on practical test sets with an F1-score of 0.929. In the same way, in the models for vertical link segmentation, the range of F1-scores was from 0.799 to 0.932 at different character positions, which proved the robustness and precision of their methodology.

The segmentation techniques that Xie et al. [29] describe in detail are valuable in accurately recognising Chinese symbols. However, the text cites segmentation as a vital repetitive process in character recognition, which involves marking the boundaries of text lines and individual letters.

The technique involves collecting pixel information across a row and then finding an average, which results in a column vector. The resultant vector differentiates between lines with content and blank lines, making accurate text line segmentation possible. This type of segmentation is needed to uphold the structural features of the characters during their recognition.

After segmenting text lines, the next task, as outlined in the paper, is defining the approach used in letter segmentation. This can be achieved by projecting an image downwards to form a row vector, highlighting the spaces between the characters. Such empty spaces act as markers for segmentation, such that the starting and ending points of the characters are used to mark Chinese characters. The technique ensures each character maintains its features, and yet it can be distinguished from the remaining ones. The proposed segmentation approach is robust since it is combined with a Convolutional Neural Network (CNN) such as AlexNet.

Segmentation of Arabic Handwritten Characters (AHCS) is characterized by unique difficulties due to the ligatured and contextual nature of the Arabic script. The existing segmentation approaches generally utilize Segmentation Points (SPs) to segment the characters into distinct characters. However, in this context, El Khayati et al. [30] propose a new characteristic diagram that alters the segmentation approach by presenting areas instead of Segmentation Areas as SA in place of normal fixed points. This modification helps to overcome major drawbacks generally associated with standing SPs. SAs are latitude areas, all the pixels located between two immediate characters, which form the boundaries proposed by the Ground Truthing Method (GTM), which is a more detailed interpretation of boundaries. This is highly relevant to the Arabic script, where several options for SPs are relevant due to the cursive and overlapping nature of the writing. Two distinct types of SAs are defined:

Normal SAs: Borders between two separated characters aligned horizontally.

Overlapping SAs: Borders where two or more characters overlap or even touch, along with vertical structures, while having more segmentation points.

The GTM dataset generated with the help of this new framework is an important contribution for the researchers.

The dataset contains over 11,000 characters and 10,000 segmentation areas, all working together to help assess the performance of AHCS algorithms in the best possible way. The SA-based derived method is robust and allows for many forms of character spacing and overlapping aspects, which are standard in handwritten Arabic script.

Huang et al. [31] proposed a new technique of character recognition from multi-style vehicle registered plates using the improvements in deep learning algorithms. Their approach uses Mask R-CNN which is an advanced technique for instance segmentation, to solve problems such as overlapping characters, font diversity and dirty background associated with license plate character recognition.

Character segmentation using contour or edge detection is the most commonly used approach among segmentation algorithms. These techniques have serious problems with overlapped or blocked characters, plates with many lines, and plates with different-sized and styled characters.

Here each character is treated as a different object, and with the plate being one different object allows for easier differentiation between multi-line and variable length plates.

The architecture's key point is implementing a Feature Pyramid Network (FPN) that generates a number of semantic feature maps while only using a single input. With this architecture, the model is able to effectively locate and segment small objects, such as individual characters, as well as large objects, such as the whole license plate. Dealing with issues created by differences in the size, font, and style of characters is ameliorated by the fusion of high-level context with low-level detail using the FPN.

Using the instance segmentation approaches background clutter is no longer an issue. Besides, the bounding box and mask predictions produced by Mask R-CNN permit to correctly distinguish elements such as characters on a license plate from the background. In addition, the approach contains an assembly method which combines bounding box and the mask information to reconstruct all the letters in license plates out of the characters. This guarantees greater accuracy for cases when parts of plates are covered or when plates are stacked.

Automatic Number-Plate Recognition (ANPR) is one of the components of the Intelligent Transportation Systems (ITS), which involves segmentation and character recognition techniques with high reliability. Wen et al. [32] employ a new approach to increase the efficacy rate of the LPR system with the use of more advanced segmentation and post-recognition techniques. Their approach focuses first on addressing tilt and distortion problems that are often barriers to correct interpretation in real-world applications.

The segmentation method in the Work begins with the detection and normalization of the plates to a fixed size of 100 × 200 pixels. This fixed normalized scale ensures consistency in the subsequent phases of the process. The authors identify two types of corrections in their method since license plate photographs are sometimes skewed both horizontally and vertically. In the beginning, efforts are concentrated on obtaining a horizontal adjustment to account for the misalignments of multiple large characters. The method employs a connected components approach to detect the various characters on the plate and estimate their angles of rotation. After that, the average angle of rotation is used to pivot the image using a 2D rotation method. This makes sure that the characters are perfectly aligned to the horizontal plane, which improves segmentation and recognition accuracy. However, even after horizontal adjustment, vertical angles might still impact the smaller characters present on the plate. The vertical rotation angle is determined by finding the minimum projection value, and this method is explained in detail in a preceding section. After that, horizontally shifting pixels corrects for the vertical orientation, allowing for better distinction of the small characters because the numbers are arranged so closely together.

Segmentation of characters is one of the most crucial elements of efficient systems of recognition of car dashboards and car dashboard recognition systems. Gao et al [33] worked out a method based on contour analysis for better accuracy in locating the pointer and digits found on car dashboards. Their approach combines advanced techniques in image processing with applications of machine learning, and as a result, it is highly accurate and robust over a wide range of dashboard variants.

Contour analysis is, and Wong et al [40][41][42] focus on traditional geometric modifications (rotation, mirroring, cropping, and a thinning approach is employed to modify the connected component of the pointer to a one-dimensional skeleton. This makes it feasible to obtain a good approximation of the position of the pointer. This methodology is helpful in the research of dashboard indicators, where the position of the pointer and accurate fixation are critical.

The authors state that a characteristic is distinguished by an SVM classifier, which is integrated with an HOG descriptor. This combination takes advantage of HOG's capacity to accumulate shape details and SVM's capability to perform classification. The approach is implemented in stages such that it enhances the efficiency of the algorithm in the differentiation and recognition of dashboard characters. The staged approach also ensures that even in the worst-case scenarios, character recognition mistakes are greatly mitigated and accurate fine-grained segmentation takes place.

The algorithm was rigorously tested against alternative approaches using a dataset consisting of 500 images of car dashboards. The performance evaluation of the algorithm was done on the car dashboards within a controlled environment, which enabled all comparable data to have proper benchmarks. The success metrics indicated the high precision of the algorithm as well as the scalability across various dashboard aesthetics which is supportive to the practical utilization of the algorithm.

Girdhar et al [34] have conducted a detailed review of segmentation methods that form an integral part of studying the historical documents. The authors also expand on techniques for text segmentation which are necessary for dealing with scanned relics, in addition to that, they also propose new ways to overcome the challenges posed by the complex format, cursive texts, and superimposed characters.

Projection Profile Based Iterative Statistical Block Analysis: It is used for determining text baselines contained in images of handwritten scripts.

Fuzzy C-means Clustering: This method seeks to overcome the problems experienced in the extraction of text lines from overlapping cursive and multi-touch text. It works by similar grouping and thus fully implements text line separation, which is valuable for reading and other processing activities.

DBSCAN Algorithm: A method that employs density clustering within a scattered environment to help with line segmentation without the need for triage. This approach is particularly effective in working with complicated layouts since it does not rely on traditional binarization that can affect the textual content of historical documents.

There is the use of Fully Convolutional Networks (FCN) that help slice the pages into important sections. The way the FCN is architecture with convolutional and deconvolutional suffices to partition the content of historical pages into zones which enhance the clarity of the document and the process.

The research focuses more on the application of current deep learning algorithms, as character segmentation is regarded as a sequence labelling task. This paradigm shift makes it easier to perform accurate recognition and segmentation of characters in the presence of the challenges of irregular handwriting and decayed scripts in historical manuscripts.

The mixture of languages and variations in writing styles and structure along with script characteristics make the task of segmenting bilingual handwritten documents a complicated one. According to Kaur et al. [35], Gurumukhi and

Latin script documents also share similar complexities and add in some classical techniques along with a heuristic based segmenting approach they suggest a solution for such documents.

The alleged segmentation methodology claims to combine projection profiles with the statistical and structural features of text which is said to classical and improves the quality of the segmented documents. This hybrid scheme is adept for other elements such as directional language segmentation difficulties, especially for handwritten documents where variability rate is very high.

For word segmentation, the authors propose an algorithm for end-point detection that detects words through gaps inside the words. This is especially important for the area of handwritten text, which may not always have clear boundaries between words. This algorithm also guarantees the accurate segmentation of words, hence increasing the reliability of the system.

Once every segmentation is done, a post-processing step is performed to enhance the accuracy and eliminate errors. The researchers tested and evaluated the method extensively using bilingual Gurumukhi and Latin datasets. The effectiveness of the approach is satisfactory as seen by the segmentation results as the average accuracy of text line segmentation is 92.95% and the word segmentation is 92.25%.

In [36] Han et al explains a method for an individual character is extracted using a pixel projection-based picture to binary conversion. All of the characters in the segmentation are captured within a minimum axis-aligned bounding box defined in terms of its location coordinates, width and height.

Suthar and Thakkar [37] delineated hierarchical segmentation of handwritten Gujarati characters for dataset creation handout. Their methodology employs three strategies: Line segmentation, word segmentation, and character segmentation; with projection profile techniques for each to aid in efficient text extraction.

The creation of horizontal projection profiles enabled the authors to draw line boundaries around handwritten images. Vertical projection profiles were employed so as to divide lines into words. To create a column histogram, the word counting method recorded the number of black pixels located in each column as it scanned the word image vertically. Characters within the scanned word were further scanned using the vertical profiles to determine word boundaries. Individual characters were defined using distance measurements that were thresholded for purposes of discrimination.

Zhang et al. [38] present a novel approach for the segmentation and restoration of stone inscription characters.

The system contains two GANs, known as GAN S1 and GAN S2, each specializing in a separate role. S1's role is to recognize the standard printed characters and also generate them, and to a smart decoder, who reconstructs characters that are partially complete. GAN S2 utilizes the outputs of GAN S1 but focuses on the further improvement of segmentation and restoration of the stone inscription characters. This is done by specifying character masks and using the learned parameters from Gan S1, thus enabling better segmentation of the inscriptions that have various forms of damage. The effectiveness of the framework was demonstrated through two datasets: a real stone inscription dataset containing 9,000 items and a supplement consisting of 90,000 standard printed characters.

2.3 DATA AUGMENTATION TECHNIQUES

2.3.1 INTRODUCTION

Data augmentation techniques that enhance the diversity of training datasets, without impairing the attributes of the script, remain some of the main issues within this domain. Rotations and geometric alterations are among such adjustments, and have been effective in expanding datasets. A primary limitation of these approaches is that they tend to make global changes that may obscure the crucial detail needed for distinguishing complex scripts. For this reason, more targeted variants have emerged that address these gaps.

In this literature review, the authors carefully evaluate the progress of various data augmentation methods and compare them with established geometric transformations, GAN-based methods, and the recently devised stroke-based methods. Special attention is drawn towards stroke-oriented augmentation techniques, which offer a fresh perspective in enhancing the variety available in handwritten datasets by targeting the core aspects of script morphology.

The review consolidates findings from many studies and presents the advantages of the stroke-based approach for Vattezhuthu scripts, namely the ability to preserve visual and stylistic details of the characters while solving the problem of recognizing ancient handwritten scripts.

2.3.2 REVIEWS ON DATA AUGMENTATION TECHNIQUES

In data augmentation processes involving Khalifa et al., Chowdhury et al. and Wong et al [40][41][42] focus on traditional geometric modifications (rotation, mirroring, cropping and scaling). These are effective techniques; however, they often change the image globally. The proposed stroke-based augmentation approach is different since it captures certain features of the handwriting style of some letters in the Vattezhuthu script, hence heightening diversity without altering the primary designs of the letters. As it applies to specific characters only, this method is advantageous for systems dealing with complex handwriting, unlike the usual methods that apply to single character systems.

Hayashi et al. [43] and Antoniou et al. [44], on the other hand, explain that there is a possibility of deep learning neural networks, Generative Adversarial Networks (GANs), to be used to form several handwriting styles that have not been used before. These methods involve completely new data; however, the method proposed works on previously existing data and alters features on the stroke level. The proposed method minimizes the variations of strokes so that the target character's stroke structure remains intact, enabling minimal boosts that are realistic compared to typical aggravations, which tend to create unrealistic data, a challenge in GAN-based approaches.

As noted by Mahendran [45], for Japanese characters, the enhancement can be done through strokes by making some lines thicker or thinner or even longer. The proposed Vattezhuthu method follows a similar route, but it is more sophisticated and has a component that is based on directionality. This allows one to be more accurate, even by the way they look at and change the direction of strokes, such as lines and curves. Such a level of detail helps maintain the organic form of handwriting, which, of course, is important in its use in scripts like Vattezhuthu.

Tabassum et al. [46] point out that augmenting alters the recognition accuracy, and it improves in particular cases of smaller datasets. The proposed way is aimed at the same target of improving performance, but this one sub-targets the improvement of specific features of the script. So far, most of the studies that have been conducted employed general transformations such as rotation or zoom. This method, on the other hand, interprets the script from the view of unique strokes of Vattezhuthu, which is a very productive addition to the shape of the script.

Similarly, just as in the Vattezhuthu case, Joseph et al. [47] do emphasize however the need to augment scripts that are rarely utilized such as MODI. But the proposed approach goes farther by seeking structured, stroke-oriented enhancements rather than simple changes. By attending to particular strokes, the character's appearance is preserved as it was intended by the original designer, which is being reinvigorated by GAN-based techniques do have ability for innovative sample generation, however, in their uncontrolled mode they create noisy or distorted data. The approach suggested strengthens control over character augmentation through management of directionality-based stroke level erosion and dilatation ensuring that the samples are authentic in the style of handwriting produced.

As specialized methods become more and more important, Wen et al. [49] and Tabassum et al. [46] recommend stroke-based augmentations for Chinese and Arabic characters. This offers a new method of advancing the study of Vattezhuthu because it resolves stroke level problems while bringing in more data to train on but does so with care to the visual presentation of the script.

The techniques for rotation, shifting, and augmenting colouring have been previously employed by Nasim et al. [50], Jungiewicz et al. [51], Arthi et al. [52], Islam [53], and Andrade [54]. Such tools, however, although effective in enhancing the scope of the learning set, are devoid of intricacy, which is perfect for scripts that are handwritten and hard to decipher. This is unlike other broader approaches for data expansion. That said, one of the scanning approaches used in Vattezhuthu is, however, inefficient in the interpretation of Chinese handwriting.

In Named Entity Recognition (NER), Islam [53] speech incorporation of text does not correspond to graphic augmentations of script characters since the method seems to focus on improving word embeddings. In a like manner, Andrade [54] works on oranges and employs primary picture processing features such as zoom and rotation. Despite being effective in its own discipline, it stresses the difference in complexity levels because dealing with the morphology of ancient scripts requires more precise and context-sensitive augmentations using the Vattezhuthu script rather.

On the other hand, the stroke-based augmentation method proposed is more accurate and less delicate than both the traditional and deep learning-based methods. As a result, it is more suitable for old handwritten scripts, which are being reinvigorated, such as Vattezhuthu.

2.4 CLASSIFICATION TECHNIQUES

2.4.1 INTRODUCTION

This section provides a comprehensive review of existing methodologies and techniques employed for recognizing Indian and foreign scripts, offering a comparative perspective on their effectiveness and limitations. The discussion is wrapped up with a table wherein the various systems in literature are evaluated in terms of accuracy and the changes that learning systems have undergone. The following subsections explain the important stages in the character recognition process, outline some related works, and prepare for the construction of stable solutions to the problems of recognizing ancient Vattezhuthu scripts using visual transformers.

One of the challenging tasks in machine perception is concept recognition of handwritten characters, as it tries to imitate the ability of a human being to decipher a written script. With the increased access for many to ancient documents, the preservation and retrieval of information in these documents have become crucial, with mechanization being the answer to the mental stress of coping with document preservation and physical typing. By using learning-based pattern recognition systems, they were able to come up with automatic character recognition systems that not only make the processing of historical data faster but also more effective.

On Indian and foreign scripts, character recognition algorithms and their various stages, including the pre-processing stage, feature extraction, and classification techniques, have been widely investigated. Nevertheless, Vattezhuthu is an example of a micrograph that not only has complicated structures, but there are also no standard datasets available, plus its source materials for the recognition process are in an extensively damaged state. Vision transformers (ViTs) aim to work out complex spatial dependencies and contextual information, which makes them a potential supplement for classical convolutional neural networks to be applied in offline handwritten character recognition.

2.4.2 CLASSIFICATION BASED REVIEW

In this paper, Maheswari et al. [55] describe an intelligent character segmentation and character recognition system to read manuscripts, which they intend to automate to create a reading system for archaeological and historical manuscripts. The scans of palm leaf manuscripts, which include Agasthiyar Vaithiya Kaviyam, Ramayanam, and Thiruvilayadal, are fed into the system, followed by preparing the data to remove noise, crop the data, and remove punch holes. A new approach using a three-step segmentation technique solves the problem of overlapped characters formed in a row. Also, a gap at which a segmentation case occurs, which was not solved by the done techniques, was solved by the proposed Augmented Horizontal Projection Profile (AHPP). Also, the algorithm for removing a punch hole and the cropping approach for containing the content within a box within an image eliminates a number of efforts needed in real time. An architecture with Convolutional Neural Networks (CNN) that has been trained with 125 classes, so that all 247 letters plus the 12 numerals available in the Tamil language can be recognized with low complexity in the network, has been used to recognize Tamil characters. The average values of the proposed approach are 98.25 percent for the segmentation accuracy ratio, 96.04 percent for the recognition accuracy ratio, but only 0.21 percent for the loss ratio. It should be noted that the proposed method outperformed existing feature extraction methods and any other CNN variations. Further developments include a model that will convert the application into hardware for Tamil letter recognition for manuscripts made on palm leaves, and the recognition of manuscripts for other languages, which will help to study history better.

The paper by Sujata Saini et al [56] presents the dataset of handwritten Indus signs as reported in the inside pages of this paper, which was developed through a social participatory approach and is entirely new. The Indus script is a writing system that has not been broken, despite many attempts, due to the lack of data and evidence that has the potential to support the needed material. There is a broader and wider manpower and material cost in bypassing the naive way of applying advanced machine learning techniques. This work aims to mitigate the challenge of scarce images of genuine Indus signs by creating a web-based application to collect handwritten replicas. To this end, 44 participants made ten Indus signs on top of the original signs, and here’s a new dataset. Applying this dataset, convolutional neural networks demonstrated an accuracy of 70% in classifying images of original Indus script writings.

According to Elshehaby et al. [57], the automated process of recognizing cuneiform symbols has been taken a step further with the use of deep learning techniques. With the VGG16 model, the authors constructed a Cuneiform Symbols Recognition System (CSRS) capable of identifying symbols appearing in the Code of Hammurabi, confirming the possibilities of transfer learning in reading ancient scripts. This endeavour is within a larger movement that seeks to apply artificial intelligence tools in investigating cuneiform tablets.

The methods employed in Modi script character recognition include a variety of approaches that have changed with time, contributing to the development of recognition technologies regarding this ancient script. The Optical Character Recognition system (OCR) has been the most basic system, using various methods such as Convolutional Neural Network (CNN), Artificial Neural Network, Support Vector Machine, Chain Code representation, and Zeroth-order moment feature extraction for its implementation. The impact of deep learning has also contributed to improved recognition of the writing system by considering the modifications made to the structure and the style of the script.

Deep learning methods have excelled in handwritten Chinese character recognition in recent years. Yet, limited attention has been given to rare characters from Chinese classics, which are challenging to input due to their numerous strokes, complex features, and often unknown readings. The authors Y. Gong et al [58] explain that a literature critique reveals that many works remain in relatively under-explored fields, which this work aims to address. This paper addresses this gap by constructing a dataset of 290 rare Chinese character categories, comprising 71,277 labelled samples extracted from the Guofeng chapter of the Book of Songs. Unlike the CASIA-HWDB dataset, this dataset is more specialized, targeting uncommon characters and serving as a complementary resource. Several approaches have been deployed to implement the recognition task, such as the Convolutional Neural Networks (CNNs) approaches; however, CNNs approaches have limitations regarding data augmentation, model lightweighting, and robustness. These will be harder in languages with a national pride and complex writing systems. The handwritten Japanese characters are supplied with immense amounts of data, without which, creative content is difficult to collect with a small amount of data. These ways have been proven productive – several fully automated systems with a developed recognition and translation ability exist today, which achieved a classification accuracy of 100%, a recall of 99.89%, and the mAP of 99.6%, while single character recognition accuracy reached 99%.

Graham West et al [59] describe a new approach in deep learning, which is meant to automate determining the date of ancient Greek papyrus fragments using images only. The pipeline elaborates on four different processes, like handwritten text recognition (HTR), which detects and classifies the characters, filters and groups the characters, and creates date prediction models for each character and each component. A brief large corpus was developed for this study, consisting of around 7000 fragments and about 770,000-character images obtained from such digests and then passed through HTR models. The authors applied transfer learning and used it to the ResNet architecture in predicting the characters' level timestamps, although they had low accuracies ranging from 35%-45%. Nonetheless, it assisted in building a fragment-based model with 79% accuracy in predicting the date of the pieces that had sufficient information, and the pieces encompassed a 2-century gap. In conclusion, the paper provides an insight into the usefulness of deep learning in the analysis of manuscripts and their dating, while at the same time appreciating the limitations in the accuracy of date prediction and the need for additional proof on other image sets.

To improve the recognition rate of Tamil handwritten characters, C. Varshini et al [60] use the ResTANet model techniques. First, the Mendeley website obtained some datasets concentrating on Tamil documents. Gaussian filtering and Otsu thresholding techniques were used to denoise the data and perform binarization to improve the data quality. The backbone of the model is developed based on the ResNet101 architecture, which has been suited to the model's needs to learn the necessary features from the Tamil handwritten characters. Also, the model employs Adam optimizer, one of the most widely used optimization methods for deep learning in training models, in optimizing parameters such as the characters' length, width, and various statistical attributes, among others. The performance of ResTANet was thoroughly assessed, giving an accuracy of 99.37% as compared to the performance of numerous contemporary algorithms in the field of handwritten character recognition, an exceptional result. The high level of accuracy achieved in the recognition of Tamil handwritten characters shows how efficient the methods used in the ResTANet model are. It thus marks a great effort in the encoding and archiving ancient Tamil written research materials.

Investigating Tamil handwritten character recognition enhancement using multimodal deep learning techniques is a step forward. In this work, R. Thendral et al [61] tackle the problems posed by the complexity of Tamil script as well as the drawbacks of previously available datasets, thus proposing the unconstrained Tamil Handwritten Character Database, which helps to boost the recognition accuracy, a key issue. The outcome of different architectures of deep learning used, such as RESNET50, VGG16, and LeNet50, has contributed significantly to feature extraction and classification methods. The large-scale tests conducted demonstrate encouraging increases in the recognition accuracy, which opens new opportunities for further work in the field of multi-model learning and in the context of preservation and digitization of Tamil handwritten documents.

The study by S. Devendiran et al. [62] recognizes Tamil characters as a great challenge, and handwritten ones are especially challenging to identify. The intricacy and form of the script, which is cursive and has a lot of diacritical marks, make automatic recognition of the script quite a daunting task. This paper explicitly focuses on handwritten Tamil characters and their optical character recognition (OCR) using two deep learning models. They employ intelligent feature extraction strategies in the two models. Model A employs the VGG-19 architecture while Model B employs Histogram of Oriented Gradients (HOG). For this research, they have compiled a wide array of nearly one thousand handwriting styles by collecting samples from a diverse population spanning all age groups. Using a multistage strategy, they prepare the images to remove any obstructions to maximize feature extraction, such as cropping, resizing the images, noise reduction, and colour conversion. The study aims to investigate the effectiveness of the VGG-19 based convolutional neural network (CNN) in addressing the complexity of Tamil character recognition compared to HOG as feature extraction. There is a considerable increase in the accuracy of optical character recognition (OCR) of Tamil script and its reliability indicating a potential application in converting paper based old manuscripts to electronic texts, digital teaching, and saving the ancient language.

HCR is an important and multifaceted area of research and practice with significant implications. In a world where the reliance on digital data is ever-increasing, HCR has a number of achievements and findings to show. HCR is used in many sectors, such as the archival digitization of historical texts and cursive scripts, self-addressed envelopes, form filling, and helping disabled persons through voice synthesis of handwritten notes. Recognition of handwritten text poses enormous challenges due to the wide variability in individuals' writing styles, letter shapes, and noise resulting from poor writing or scanning conditions.

It is necessary to note that recognition systems ought to adapt to this diversity. The progress in deep learning techniques, particularly convolutional neural networks and recurrent neural networks, has dramatically enhanced the performance of handwritten character recognition systems. These models can detect complex relationships and associations in the handwritten data. In this work, Sunitha S. Nair et al [63] describe a relatively efficient HCR model based on CNN-RNN with Bi-GRU networks. A well-structured handwritten character dataset has been developed. The accuracy obtained by the proposed model is equal to 96.72%. The CNN part is essential in processing spatial features and information about the shapes and structures of the characters from the input images. The RNN with Bi-GRU layers performs an important synthesis of sequential structures that are needed in recognizing the temporal attributes of the handwriting, especially when characters are written in a more or less connected mode or when the characters are written in different ways. This combination of techniques allowed achieving automation, enhanced data processing, and improved user experience in various fields.

Handwritten text recognition (HTR) plays a very important role in computer vision because it has far-reaching consequences for many different fields. Due to the growth of deep learning and machine learning algorithms and more raw data that is not formatted, many researchers have gained a footing in HTR. In this work, Supriya Mahadevkar et al [64] aim to bridge the gaps in HTR by using a different approach to submitting findings. The specific focus of this was to achieve a higher degree of accuracy in reading the photograph's length. Results show that integrating CNN and BiLSTM layers with a CTC decoder drastically improved the precision level. As experienced practitioners of the IAM and RIMES datasets commendably attained the accuracy of 98.50% and 98.80% respectively.

In its most fundamental explanation, OCR is a technology that concerns using textual data obtained from a scanned image or photograph. This research proposes a new approach to optical character recognition (OCR) using the VGG19 CNN architecture. VGG19, famed for efficiency and depth regarding picture classification, is also used here to solve the complicated problem of recognizing characters. In this work, G. Singh et al [65] demonstrate the efficacy of the proposed approach in obtaining state-of-the-art accuracy and resilience in extracting text from images after going through thorough experimental and authenticating processes. This research provides randomized datasets of printed and handwritten text samples and enhances them via augmentations for increased model generalization. The VGG19 model is trained end-to-end, treating its convolutional layers as feature maps for character recognition. This work presents a new approach for OCR based on VGG19 convolutional neural networks. Firstly, OCR is a key technology that allows us to convert printed or handwritten documents into notation and therefore allows for the digitization and search of material. The proposed method employs hierarchical features that use VGG19 to focus on the layers to grab the textual information from the photo.

In this study, experiments were carried out using publicly available datasets, and the goal of this work was to reach higher training and test accuracy across epochs. An accuracy of 94.34% in training and 94.96% in testing was achieved by the proposed model after 10 epochs. Additionally, they observed a persistent decrease in both training and test loss over the training phase, indicating proper convergence and optimization of the model parameters. The findings demonstrate that the VGG19-based OCR model is competent in recognizing and retrieving textual information from diverse digital images, which opens a wide range of applications in practice such as digitization of documents, augmented reality, and assistive technologies.

Prof. Jyoti et al [66] have outlined a constructive approach involving the integration of convolutional neural networks (CNN) and bidirectional LSTM models to enhance the accuracy of handwriting identification. It captures the possible information that both models possess when they work together to augment their models' detection abilities. The results of the experiment provided evidence for the effectiveness of the fusion technique since it was able to outperform both LSTM and CNN models as well as across two datasets: IAM and RIMES. The first step of this research was to train the two models, LSTM and CNN, on the IAM dataset. The model, based on long short-term memory networks (LSTM), used a bidirectional architecture to gather information about the environment in both directions, i.e., from the past and from the future. In contrast, the LSTM model took the input images to get significant environmental features. The long short-term memory (LSTM) model achieved 89.5% accuracy, while the convolutional neural networks (CNN) accuracy is 88.2%. The results also provided evidence for the robustness of the two models when applied to the task of individual handwriting recognition.

Prototype Role and Access Management RIMES Accuracy Strengths MPPT analysis providing validation accuracy LSTM: 89.5%, 82.3% – Mostly assumes the context of previous/next CNN feature vectors 88.2% and 79.6% – Extract robust local features Overview of Proposed Fusion Model: 91.2%, 84.7% – Expands the operative scopes of both LSTM and CNN, guarantees quick optimization, is noise and occlusions resistant and applicable for various fonts styles. After the LSTM and CNN models have been separately developed, they are combined through the late fusion approach. In terms of model output, both models were combined using a weighted averaging method. The fusion model achieved an accuracy score of 91.2% exceeding any of the individual models' scores. This enhancement can be attributed to the combination of local information obtained via LSTM and local features obtained through CNN, which broadens the scope of the article being considered. This means that a variety of deep learning approaches could be applied together in order to get the best out of handwriting recognition. A validated manuscript fusion model adds credence to editing the RIMES data set, which uses complex manuscript models. The LSTM and CNN models on this data set achieved scores of 82.3% and 79.6%, respectively.

The fusion model yielded 84.7% accuracy, which represents a substantial improvement over each of its individual counterparts. This corroborates the fact that the fusion model is able to appropriately cope with the complexities that characterize the real-world manuscripts and is able to generalize well on the different data sets. Besides improving the accuracy, the pictures' noise and binding effect did not affect the proposed fusion model much. The LSTM model can collect distance information, while the CNN model gathers information about the nearby people. The performance of the fusion model did not change with an increase or decrease of noise or occlusion, which indicates it could perform in varying environments. It has been noted that the classification algorithm works faster when using just the LSTM model, if compared with the LSTM and CNN simultaneously. Greater collaboration using the two models can only make it more robust. The consensus model, it would appear, cut down on excess use of computational power and training time and managed to achieve convergence within a short span of time. The impact of the fusion model was also studied in terms of how well it performed on other document types, such as lowercase, uppercase, digits, and symbols. It is also worth noting that across all the categories of documents, the fusion model once again proved to be far better than the individual models.

This all-in-one crossbreed model makes it applicable and useful across several other sectors relating to the field of penmanship which includes document scanning, document digitalization as well as automated document recognition. Always, however, this guideline has enabled me to make several inferences in the context of future developments in this area as well. This study looks at several aspects of this analysis, such as the various fusion mechanisms. This study used an integration approach whereby the outputs of the models based on LSTM and the weights of CNN were computed and combined. More procedures, such as starting fusion and later fusion, might englobe an enhancement of the efficiency of structural fusion. An additional realm of potential research is the range of educational data. Enhanced performance would surely have been achieved if there were enhanced datasets available for training the fusion model, for instance, the IAM and RIMES datasets that were utilized in this research, even though they are rich and comprehensive. Furthermore, extra training data would improve accuracy for the parameters that determine the performance of the fusion model due to its ability to ensure correctness in output, considering varying letters. In addition to those two studies, further research might also focus on evaluating the impact of both LSTM and CNN models based on the specific architectures. Choosing appropriate hyperparameters and network structures significantly affects how effective deep learning models can be. More work on other structural properties may also improve handwriting text recognition.

This research proposed a fusion of Bidirectional LSTM and CNN models to enhance handwriting recognition. The seamless integration of models has been noted to do much better than the respective models on the benchmark datasets and has shown good resistance to noise and occlusions. The fusion model showcased such great versatility that it was broadly relevant across many manuscript types. Future work should test other fusion methods, increase the training dataset, and try different network architectures to improve the performance of the proposed model.

The paper [67] incorporates an Adaptive Convolutional Neural Network (ACNN) specifically for the Tamil Handwritten Character Recognition (THCR) task that creatively transforms its convolutional filters during the training session in order to cater to the many and diverse forms of Tamil handwritten characters. With this ACNN system, the model is able to capture and absorb handwriting styles from various regions, leading to a higher accuracy rate. The experiments show that the ACNN is better than typical convolutional neural networks, which suggests that the suitability and the performance of the ACNN in recognizing Tamil characters have improved, and hence, a viable system for digitizing and archiving the rich Tamil cultural and linguistic assets.

This work [68] studies the use of both deep learning and traditional machine learning approaches for HCR. Convolutional neural networks and hybrid models were used for Arabic, Hindi, and Latin handwriting recognition. The accuracy of models trained on the MNIST dataset was 99.31% for training and 98.4% during testing, which indicates overfitting was unlikely to occur and generalization was present. A 99.45% accuracy was achieved, whereby the convolutional neural network (CNN) integrated with the bi-directional long short-term memory (Bi-LSTM) surpassed the conventional CNNs in cursive script recognition. Cursive dependent characters and related characters were detected by Bi-LSTM as it was able to understand the sequential dependency as well. Using such pre-trained networks as VGG16 or ResNet50 proved beneficial in terms of accuracy, with ResNet50 reaching an accuracy of 99.67%, thus demonstrating that models can be effective when used with little amounts of training data. The experimental results of single-shot models using RNNs and LSTMs showed effective model performance, and there was also an obtained character recognition level accuracy along with the scripted problems. Attention processes of different visual characters, such as '3' and '5' or '0' and 'O ', were highlighted as hybrid models managed to reduce their misclassified instances and recognized the need for further refinements. For the models, most of the Latin characters were easy to deal with, while grappling with the Arabic scripts, which were described as the diacritical marks, and the ligatures were the hardest. It was observed that there were some attention techniques employed as well as the segmentation improvement approaches, which substantially contributed to the performance of the model in terms of the accuracy of self-embedding the different scripts. The work endorsed CNNs, hybrid models, and transfer learning, which would serve to improve recognition systems in terms of accuracy and versatility. It suggests that more effective HCR systems can be achieved through the use of transformer architectures and more universal models. Traditional techniques can be complemented with hybrid approaches.

In this research work, Humayun et al [69] focus only on one language, English. In particular, it uses offline handwritten character recognition, which aims to read/write English letters. Many competing datasets are readily available, the most challenging being EMNIST. This research work proposes a deep learning-based ELBP-CNN approach for recognizing English characters. This research work describes a deep learning CovNet that fuses feature extraction and local binary pattern-based methods, LBP (AND, OR), that are assessed and compared with mainstream pre-trained models using transfer learning. This approach focuses on empirical experiments with models and looks at how certain key parameters define the models. All models tested in this study, as well as E-Character, involved the same hyperparametric configurations across models, irrespective of data augmentation parameters. The model developed and named E-Character recognizer produced an accuracy level of 87.31%. Outperformed most of the pre-trained models and the other methods offered by other researchers, which were evaluated.

This study [70] gives an approach towards implementing the digitization of Malayalam PLMs using an ensemble which combines CNN and BiLSTM networks which allows us to automate Malayalam PLMs. This technology could potentially serve as an important tool all around the world for medicine, research, history, astrology or archaeology. In the age of the COVID-19 pandemic, when the world is battling to contain the spread of new viruses, digitising Malayalam PLMs that contain different Ayurvedic recipes is very much needed.

As outlined by our literature review, there is a clear gap in this area that warrants further research. Thus, our efforts are going to greatly assist in the survival of this priceless information for the generations to come. The model therefore achieves a fine-tuned level of 96.40 percent, greater than the existing best practices. Despite the similarity in the appearance of the manuscripts which were analysed from 1800 to 1908, the Malayalam font has only changed gradually through the ages.

Historical handwritten documents are prone to degradation, which requires Optical Character Recognition (OCR) technology for efficiency purposes. The complexity of recognizing characters is compounded by several aspects of ancient manuscripts, such as humidity marks, faded text, overused ink, or text that consists of similar letters. In this paper, Jindal, A et al [71] present an optical character recognition (OCR) technique utilizing an efficient feature extraction network based on the convolutional neural networks (CNN) that aims to automatically identify all characters inscribed in the ancient text written in Grantha. The new method uses several layers of convolutional layers for a deep hierarchical feature vector representation of the input character image. Two fully connected neural network (FCNN) layers have correctly classified these feature vectors into many of their classes. The hyperparameters of the CNN model used, which include the number of filters per convolution layer, filter size in each convolution layer, the number of FCNN layers, and neurons in each FCNN layer, were calibrated using the Bayesian optimization technique. This work primarily presents an efficient model of the CNN for implementing optical character recognition (OCR) of ancient Grantha scripts. The Grantha OCR training system obtains a 99.30% character recognition rate for ancient Grantha handwritten text.

In this paper, Chandankhede, C et al [72] examine that it's worth mentioning that the MODI language is particularly a challenge for OCR. The average MODI script user is an OCR user who possesses a cursive style of writing with quite a few similar-looking characters. As per the review of literature, Deep learning techniques such as InceptionV3, as well as ResNet, have not been applied yet in the case of the Modi script. This promotes the possibility of using deep learning methods for offline handwritten character recognition in the structures ResNet and InceptionV3. About 25 people's samples were acquired, and this was used to create the handwritten Modi barakhadi dataset. In this paper, an experiment with a 7721 dataset is presented. For a region of interest, the individual characters are isolated, and the Otsu binarization process aids in the pre-processing stage. The analysis of supervised evaluation scores for pre-processed data is done using both algorithms on the actual handwritten characters database written by several persons. ResNet50 images achieved a testing accuracy of 94.552 % while the model achieved a precision of 0.86. For InceptionV3 images, the accuracy during testing was 93.923% whereas the model achieved a precision of 0.843.

This paper [73] proposes a new doubtless universal approach to the recognition of handwritten Devanagari characters based on entropy to manage skew states of characters and mask techniques to remove header lines, consequently reducing headers. Non-header characters assist in recovering the histograms of oriented gradients (HOG) features and are employed for non-linear projection techniques. Additional classification is also performed using the AdaBoost ensemble boosting technique, which puts the recognition performance onto a new high of 98.43% and 98.68% for the V2DMDCHAR and ISIDCHAR databases, respectively.

This work [74] presents several recently developed state-of-the-art Convolutional Neural Networks (CNNs) for automatic recognition of Bangla handwritten characters using the 'Ekush' benchmark dataset, while systematically analysing the accuracy of each CNN structure. Besides being able to learn from raw data, CNN-based systems work efficiently and effectively, which is not the case with other traditional systems. The authors of the present study illustrate that the efficiency of CNN-based models relative to the character recognition rates attains a maximum, and that means CNN architectures are reasonably effective for automating the recognition of Bangla script.

This work [75] concentrates on devising an efficient feature extraction algorithm to recognize Odia handwritten digits, standalone characters, and combined characters. Four separate approaches have been used for this character recognition. First, different characters are honoured by three distinct feature extraction processes, and later, such features are fused using several common machine learning approaches into one feature space. The second one, instead of images, provides the RNN and CNN networks with a head function on the concept of unifying features for character recognition. The same feature set and classifiers are used for the recognition of some sets of characters. The last task involved the insertion of a few selected compound

Odia characters into the proposed character recognition system. The dataset generated for such a purpose consists of numbers, basic characters, and compound characters. The proposed system achieved recognition accuracy of 86.56% for this data set, reaching 112 character classes.

The paper [76] proposes a progressive methodology for Telugu character recognition based on a multi-objective mayfly optimization with a deep learning (MOMFO-DL) model. The proposed MOMFO-DL approach incorporates the DenseNet-169 model as a feature extractor that delivers quite a useful set of feature vectors. A functional link neural network (FLNN) acts as a classifier during the recognition and printing of the characters. Designing the MOMFO technique to optimize the parameters of the DenseNet and FLNN models proves that this work is novel. With the application of the MOMFO technique, the required tuning of parameters can be achieved, and hence an enhanced performance. The extensive experimental work is performed on standard datasets, and the results are assessed according to various criteria. The experimental outcomes confirmed the efficacy of the MOMFO approach over several recent baseline approaches.

Some of India's oldest cultural documents include stone inscriptions, palm manuscripts, coins, copper plates, and much more, and significant efforts are underway to digitally preserve these documents. Implementing this can be a hassle because a lot of these images have a lot of background clutter, as well as poor lighting. Character recognition is used to analyse documents that are easily snapped. In order to gain insights from palaeographic documents, there is a requirement for a character recognition system. The process of automatic character recognition is extremely complicated as it requires the use of several technologies, including image enhancement, patch extraction, feature extraction, classification, and recognition. It involves gathering images of multiple patches and their associated feature vectors by a convolutional neural network and performing attention-based LSTM character recognition. Looking at the specifics of palaeographic texts, they have large quantities of repeated characters, which can make character recognition more difficult. S. Ezhilarasi et al [77] The primary objective of this work is to locate and demarcate Tamil palaeography of old texts by feeding a CNN-LSTM framework with altered sequences of images, which classifies the pictures into patches. The recommended approaches are image pre-processing, feature extraction, and the recognition of characters on an individual level.

The LSTM network has been built, and the series of feature vectors has been entered and trained into the network. A sequence of characters has been established. The introduced method, in terms of character identification, achieved an accuracy level of 97.9%.

The paper [78] aims to assist epigraphers by applying image enhancement methods consisting of image noise reduction, feature classification via deep learning in the context of ANCII-R-21, which stands for Ancient Tamil Character Recognition in Inscription. The pre-processing stage comprises the following steps: import of grayscale images, denoising, and binarization based on Otsu's thresholding. By merging corrosion, spots, and scratches noise, image noise is reduced by converting all noise into one single cluster. ATCRI-21 for character recognition employs three portions of a five-layer block (two convolutional layers, Rectified Linear Unit (ReLU), two max-pooling layers, dropout layer), then two two-layer block portions (dense layer, ReLU layer), and finally a softmax activation layer.

Regularization will help in overcoming data overfitting and will also allow the network to train in less time. This model was trained, validated, and tested with real images taken from Brihadeeshwar Temple located in Thanjavur. It is encouraging to note that the proposed model achieves a validation accuracy of 80%. This is the relation that will come to birth with the documentation of the history of civilization. The epigraphers' files will enhance artificial intelligence capabilities and vice versa with respect to the studied important epochs in history.

This study [79] presents a Convolutional Neural Network (CNN) model for the automated recognition and classification of characters from manuscripts, in particular, the Old English characters of the Beowulf manuscript. For the experiments in their study, they built their own character dataset sourced from the electronic photos of the manuscript and made this dataset available to all other scholars in that area of research. They built a CNN model for the purpose of performing the classification task and later on, trained and tested it using the Beowulf manuscript dataset. Also, they performed comparative analysis using other ML approaches, including Support Vector Machine, K-Nearest Neighbour, Decision Tree, Random Forest, and XGBoost. They used VGG16, Mobilenet and ResNet50 as pretrained networks to get feature representations of the Beowulf manuscript character images dataset, used these features to fit the SVM, KNN, DT, RF and XGBoost models, and later tested these models on Beowulf test dataset. The recognition accuracy results were presented, and for performance evaluation of each model, recall, precision, and F1 scores were calculated.

The ROC curves were supplied in order to subjectively assess the classification ability of the CNN model relative to the other models. The dataset of the character pictures from the Beowulf manuscript was first, second, and third time augmented (that is, resampled) once, twice, and thrice, respectively. ML models were outperformed by the CNN model, that they are recommended, with the CNN model recording the highest accuracy of 98.86% when augmented resampling 3. In addition, the CNN model has also been trained and analysed on the MNIST dataset for handwritten digits recognition, where the standard for recognition accuracy was 99.03%.

This work [80] focuses on ascertaining the identity of authors based on their handwriting extracted from papyri among other archaeological sources, with regard to the numerous challenges posed by the potential of archival historical documents. The authors place the problem within the framework of similarity learning and implement Siamese neural networks to measure the similarity of two writing samples. The deep learning-based binarization method is employed to binarize the severely blotted out handwriting on the papyri images. Handwriting sections were created by scanning the entire picture and obtaining samples from one writer and other writers. Positive and negative pairs are then used to train a Siamese neural network based on a number of popular architectures as a skeletal model. An experimental study of the GRK-Papyri data set, which consists of 97 writing samples, managed to report a test accuracy of 62% with DenseNet-121 and 69% with a CNN, which was trained by Omniglot. The first results appear to be promising and could be applied to systems of writer identification, verification, and retrieval.

The scope of the recognition of the indentation of ancient Tamil manuscripts written in palm leaves relates to the recognition of Tamil Palm Leaf Characters. These manuscripts, which are of great cultural and historical value, require complex processing of the materials to extract characters due to the numerous chronological and degrading factors that have affected the manuscripts. In this study [81], it starts with capturing a clear picture of the document, which is first captured in a pre-scanning process to remove any blemishes or stains. Before decoding the character, the artifacts are decomposed into three parts along the vertical axis for better preprocessing results. In order to remove the issues with colour and the weight of the data, the image is said to be scaled down to black and white. Subsequently, blurring is performed, and wide-band pass filters are used. To obtain better contrast between the text and background, reduced spacing between letters and sharper edges of the letters are employed in the enhancement of the image. Binarization makes it possible to distinguish the characters from the background. The rectangular image is passed through character isolation lines in order to segment the text in the image for better recognition quality. Learning of the DCNN features of the pre-processed scripts is finally performed. The proposed model is able to recognize Tamil characters with 89% accuracy, leading to the vigilance of digital Tamil Palm manuscripts.

In the classification task, this research [82] evolves several CNNs and ensemble methodologies with the goal of enhancing the classification performance accuracy of the Oracle-MNIST dataset. Considering the projection time, the number of model parameters, as well as the average accuracy, the Com2 CNN that was postulated and snapped is the best, which performed at an accuracy of 97.007%, even better than that of Net1-3, as stated by the supplier of the dataset.

Yet such increasing integrating augmenting the training and test data of other models and employing additional capital strategies, such as combining algorithms at the graph level, should be helpful to the studies using this dataset together with Oracle-MNIST.

The history of the Achaemenid Empire was chiselled on clay tablets some two thousand five hundred years ago. An archaeological endeavour by the University of Chicago's Oriental Institute in 1933 unearthed thousands of such tablets and shards at the site of Persepolis.

Edward C et al [83] employ the dataset to develop DeepScribe, a system that seeks to delimit and recognize cuneiform signs. It is shown that a RetinaNet object detector achieves a localization mAP of 0.78 while a ResNet classifier achieves a top-5 sign classification accuracy of 89%. The authors developed a control network which during training demonstrated 80% end to end accuracy in the top-5 classification task. The sign classification module focuses on morphological sign types of cuneiform writing; these types form the basis of morphological clusters. They look at how useful this clustering technique, which is based on artificial means, can be as compared with a standard printout of a register of signs, and what difference it can make. Structures belonging to the two distinct schemas do not substantially improve the efficiency of recognition of the signs of Achaemenid-period cuneiform tablets in photographs and the automatic transliteration of such texts. Finally, they consider the sign decoding capabilities of the model in closing and outlining a proposal of a linguistically motivated transliteration system, and later try to suggest what the model might be able to do with non-Persian cuneiform periods.

The Modi Script, also referred to as the archaic Marathi script, was predominantly used to write the Marathi language prior to the 1950s. Its intricate writing style and fragmented characters have caused its meaning to be obliterated. Recently, the importance of studying older writings has increased in order to gain relevant information. scrappy images and overlapping characters in the MODI script make the documents hard to read. Since then, several classifiers have been reported in the literature, such as Decision Tree, SVM, KNN, and other hybrid systems developed for MSRS. In this study, J. Madake et al [84] develop a new Modi script recognition system that combines Zernike moments and DCT coefficients for excellent accuracy. 27 characters were defined in the created custom dataset as a class. A number of classifying methods, such as Decision Tree, Random Forest, and KNN, were used for this purpose. These other methods helped to verify the performance of the system comprehensively. Following feature fusion, a high accuracy of 97.77% was achieved.

In this paper, Ponraj et al [85] present Convolutional Neural Networks (CNNs) for Tamil handwritten character recognition. Retailing current developments, methods, and technologies, they show that CNNs can produce results with an accuracy of around 95\%. The dataset used for the experiments contains 247 Tamil Characters along with 18 North Indian characters that represent different scripts. They design custom architectures of CNNs specifically for Tamil characters and use advanced preprocessing, data augmentation, and training to enhance model performance. The work tackles the problems of the existing domain of publicly accessible datasets by providing solutions to data insufficiency, class-dominating, and writing style variation issues. The original contribution is achieving up to 95% recognition accuracy for all characters from the dataset consisting of 247 Tamil characters and 18 North Indian characters, which suggests the applicability of CNNs for the tasks of document processing, language preservation, and even automating Tamil regions. This work enhances the field by introducing new ideas, a large corpus of data, and empirical findings, representing a major progress in the recognition of Tamil scripts.

The purpose of the proposed research [86] is to improve the accuracy of optical character recognition techniques for archived ancient inscriptions in the Vatteluttu script, dated back to the 4th or 5th century A.D. Also considered is a deep learning model for the transliteration of Vatteluttu Script into modern Tamil with a prospect of expansion to other languages. All such kinds of studies are beneficial to epigraphists, archaeological researchers as well as laymen interested in this field. The constructed deep learning model reached an accuracy of up to 84.12%.

The target of this research work [87] is to be able to interpret handwritten characters and convert them into understandable text. Image rectification, segmentation, and the character itself constitute the set of characters. While the characters are input, image rectification takes place, which makes them suitable for the future processes. One of the fast-evolving areas in this field of 'computer vision' is what is called character recognition. It is only humans who have the capability of such recognition. So easily does a person identify the handwriting. For linguistic composing, different languages have a number of different patterns to observe. People will be able to read and understand the text without any difficulties. The study employs the MNIST Dataset. This strategy is based on the Convolutional Neural Network and Support Vector Machine methodology, and it is used in this case because it is accurate and performs impressively better than most of the other methods.

Yuan et al [88] commenced the development of the HABD (Houma Alliance Book Database), which is a remarkable step toward the recognition of ancient handwritten characters. The Great Book of the Houma Alliance was a problem because it was not new, and the ink layer was worn out, so the letters were barely visible. To solve this problem, the researchers collected a reasonable amount of data, which consisted of 26,732 ancient character samples of 10 categories, and detailed segmentation was done on all the samples. In addition, the authors proposed a new baseline that consisted of four classifiers based on deep neural networks augmented by two techniques that enable improved recognition of ancient text styles.

The research [89] presents a new approach to handwriting recognition of Tamil script by combining Variational Autoencoder and Convolutional Neural Networks with GAN.VAE. The study highlights the challenges faced specifically regarding HCR, such as a lack of datasets and a more language-centric focus of traditional methods. Authors were able to apply the VAE-GAN architecture to the HP Labs e TAMIL handwritten character recognizer, which resulted in a significant increase in the speed of image recognition processing. The combination of VAE with GAN provides an effective means of data augmentation because VAE compresses and reconstructs images while GAN generates realistic-looking fake data. The proposed system achieved 86% accuracy in identifying handwritten Tamil letters, showing the potential of the construct in solving handwritten character recognition problems. This study not only contributes to the fields of pattern recognition but also opens up avenues for further research in isolated character models and the use of such models in other languages.

Shai Gordin [90] constructed a specific deep learning-based OCR for Latin transcription of Akkadian which achieved results up to 89% accuracy. CuReD provides a paradigm for modifying the first resulting outputs of the model and improving the model on further unknown forms of transliteration. At least ten texts can improve the initial results of the model quite significantly.

In this paper, Savitri Jadhav et al [91] introduce a proposed classification technique that utilizes CNN-based transfer learning along with a global feature extractor, specifically the histogram of directed gradients, in order to achieve the required global invariance for MODI recognition. Moreover, the features that are invariant and give an explanation for the low recognition rate are selected using criteria obtained from principal components and the confusion matrix. The proposed classifiers are implemented using a self-designed handwritten MODI character dataset and tested on an altered MODI dataset. The results demonstrated that the framework successfully identifies handwritten characters of a MODI style without needing data enhancement or changes in the network when MODI characters undergo different transformations.

The goal of this work [92] is to design a robust feature extraction technique in order to ease the recognition of basic Odia handwritten digits, characters, and even compound characters. Four different approaches have been developed and implemented by the researchers for the recognition of the characters. To start with, three different image extracting methods are first applied on the various characters, and then the outputs are combined to obtain a common feature vector that is used in some machine learning techniques. The second method gives a more detailed definition of two well-known methods, RNN and CNN, for the recognition of the character through the provision of a well-defined feature vector, as opposed to the use of images, like in ordinary networks. There are several groups of characters that share the same feature sets and classifiers that are used to successfully differentiate them. The last task included adding a number of Odia compound characters to the proposed system. The dataset prepared for this purpose contains numbers, simple characters, and complex characters. The proposed method demonstrates a recognition rate of 86.56% for the dataset composed of 112 classes of characters.

In this paper, Gebremichael, H.T. et al [93] explain the Amharic script, which involves deep history, culture, economy, politics, religion, and science. These texts should also be shared and based on the facts, considering recent technologies like OCR, this text could be easily scanned and made available for people. This is because OCR is a process whereby it is possible to scan text documents, printed, typewritten, and handwritten, and produce an electronic copy of that document. At present, such practical applications have been created in English, one of the world's lingua franca. OCR research has been conducted in the Amharic dialect for over a decade. Recognition of Amharic-based characters began with the assignment of texture recognition algorithms and pattern recognition tasks. The goal of this research is to implement Amharic OCR for natural texts, including all the characters used in Ethiopian language texts. This paper describes the new strategies of feature extraction and provides the improved results of the Gabor Filter, followed by PCA and GA with SVM. The prototype's performance on real Ethiopic books, newspapers, and magazines was around 98.33% accuracy for the processing of Ethiopic characters.

This research [94] looks into the concerns of the CR system regarding the images of the Great Isaiah Scroll. A new dataset has been developed, consisting of a series of photographs of single letters from this scroll. They tested four CNN models on their dataset and educational activities. The results suggest that for the purpose of the recognition of isolated ancient Hebrew handwritten characters, two well-known CNN architectures, AlexNet and LeNet-5, can be the second-best and last-best models, which have a low standard deviation of accuracy and loss rate, and a testing accuracy greater than 94%.

This work [95] proposes a model for automatic recognition of Sanskrit language text characters using a 4-fold convolutional neural network (CNN) architecture on dependant scripts popular in the Indian sub-continent, for instance, the Devanagari script, which is said to have originated between the 1st and 4th centuries CE.

The MODI script, which belongs to the Nāgari family of alphabets, was used to write Marathi until the twentieth century. Although it is not used at present in an official capacity, it has some relevance historically to India because of the vast number of manuscripts available in several libraries throughout the country. This study [96] outlines the use of the wavelet transform to construct a character recognition system of the MODI script. The experiment makes use of Daubechies, Haar and Symlet wavelet, a performance comparison among these different base wavelets is done. The classification is carried out with the aid of a decision tree classifier, and it has been found that character recognition results obtained with the feature extraction method implementing Daubechies wavelet for text recognition were better.

This research [97] first applies pre-processing techniques such as noise removal, one-hot encoding, and contrast adjustment so as to improve recognition of the data set. In this study, different classical CNN model architectures are combined to create a CNN that is more flexible and more stable. This particular evolutionary CNN model consists of a coupling layer, a nonlinear sub-sampling layer, and a Dropout layer, which allows for local detail extraction while decreasing the number of parameters and data complexity. The Dense layers convert the edge image features to the output by correlating among the features within them. Each of the layers of this CNN uses an activation function based on ReLU to perform a nonlinear mapping to avoid problems such as gradient disappearance and overfitting. The estimated CNN model, Bannigidad, P et al [105] discuss the aim of this study, which is to digitize medieval final Kannada scripts through handwriting recognition. The RF, SVM, and CNN accuracies are 0.90, 0.92, and 0.96, respectively. Random Forest and Support Vector Machine make some progress in the process of finding several empires of the characters, that is, if the characters belong to the Hoysala, Vijayanagara, and real handwritten character samples were evaluated, it was reported that this CNN network obtained the best among the three-recognition accuracy and also had fewer character' misidentifications than RF and SVM classifiers.

This research includes an analysis and evaluation of this CNN against competing CNN models. Based on the experimental results, the proposed CNN provides an ideal compromise between fitting time and recognition accuracy.

R. Gayathri et al [98] highlight an exceptional method for ensuring high accuracy and high performance with regard to the handwritten Tamil character recognition system. However, the subject is still in a very early stage and, despite numerous attempts made in the field of handwritten character recognition, the accuracy level is significantly low when it comes to the Tamil language. This work applies to the recognition of the handwritten Tamil text in an offline manner and using the Inception-v3 transfer learning model. The proposed study employs the available HP Tamil handwritten character offline dataset (Hewlett-Packard Lab: hpl-tamil-iso-char-offline-1.0), which is accessible. The suitability of the transfer learning approach with the Inception v3 model permits the pretrained model to register a recognition accuracy rate of 93.1% which is higher than the other previous architectures that use deep learning. The Accuracy achieved in this work is due to the usage of a pretrained model alongside transfer learning, which reduces the time PIPC takes to achieve a certain task. This results in improved accuracy and better performance.

In this work, R. Fermanian et al [99] propose a deep character recognition network for recognizing handwritten Syriac texts. The network achieved an accuracy level of 97% in classifying the characters, which is higher than the best-known systems of Arabic handwritten character recognition.

This paper [100] describes the architecture of CNN in various forms of application: (i) building and training CNN from scratch with Softmax classifier in a sequence learning, and (ii) MobileNet architecture. Constructing a Tamizhi dataset model using transfer learning based on a pre-trained model; designing a CNN + SVM model; applying SVM in the context of the task of efficiency estimation for handwritten Brahmi characters recognition.

This research took advantage of a very large TAMIZHİ handwritten Brahmi database of 190000 isolated character samples for the training of the CNN model. The constructed dataset consists of 9 vowels, 18 consonants, and 209 classes, and classifiers of machine learning approaches can be trained. Out of all the models applied, Mobile Net proved to be effective with 68.3% accuracy level while other algorithms achieved accuracy rates in the range of 58-67% with respect to tamizhi dataset. The MobileNet model shows performance with datasets of vowels, where models were trained with eight classes, consonants, eighteen and phonetically consonant vowel syllables, twenty-six classes, and evaluation accuracies of 98.1%, 97.7% and 97.5% respectively.

In [101], an annotated database of Phoenician alphabets comprising different styles and their availability throughout time is submitted. Such a box can assist in the digitalisation of Phoenician heritage. This dataset is subjected to standard preprocessing steps followed by a hybrid type of autoencoders for processing and augmentation that still maintains the human aspect. The improved dataset is then processed with a dedicated capsule network for the recognition and categorization of the Phoenician character into one of the 22 available alphabet characters. This model surpasses other state-of-the-art models in the recognition of handwritten characters with an overall performance accuracy of 0.9891 and a loss rate of 0.021. Therefore, the models assist in achieving automated deciphering systems which save time and effort investment by epigraphists during rapid translation of Phoenician epigraphy. Furthermore, it is this model that makes it possible to study the decipherment of other ancient scripts with minimal changes since its effectiveness on Phoenician epigraphy is known. The model can provide a suitable transfer learning framework for the recognition of other developed alphabets with sparse annotations.

In this research, S. P. Deore [102] proposed the recognition of characters of the compound handwritten Devanagari letters based on the K-Nearest Neighbours (K-NN) and Support Vector Machine (SVM) algorithms. The methods employed for feature extraction from images are crucial in identifying compound characters. This work applies the Zoning technique with eccentricity, equivalent diameter, and angle as features. A new dataset of 45 compound characters has been produced. For the machine learning experiments, they performed cross-validation with a 10-fold technique and achieved a maximum accuracy of 96.62% with the SVM classifier in a training time of 1.56 seconds.

In this research, Hamid et al. [103] introduce an ancient writing technique that deserves attention: the use of Sumerian cuneiform symbols. In this work, the authors translated cuneiform symbols found in Sumerian script into the English alphabet. This purpose was accomplished with success using an MLP network. They are pleased to report that our proposed method has achieved an accuracy standard of 100%. Such aid may ease and enhance the accuracy of the explorers and researchers' understanding of Sumerian cuneiform symbols [104] applies first image processing and then deploys deep learning models with the aim of converting Ancient Tamil-Brahmi characters into readable Tamil scripts, and through image processing, enabling understanding and using the information from our ancestors.

In this paper, Bannigidad, P et al [105] discuss the aim of this study, which is to digitize medieval final Kannada scripts through handwriting recognition and to enhance the complete information by extracting HOG features. The LDA, K-nearest neighbour (K-NN), and SVM classifiers work in the process of finding several empires of the characters, that is, if the characters belong to the Hoysala, Vijayanagara, or Mysore Wodeyar dynasties. The mean percentage of some of the historical classifications of letters in Kannada handwriting under different regimes, where the LDA classifier performed best, achieving 68.4 percent, the K-NN classifier was the second best, with 85.7 percent, and the SVM was the last, reaching the lowest 87.2 percent. This research review provides that the application of the SVM classifier outperforms that of the LDA and K-NN classifiers in terms of overall classification accuracy.

In this paper Siranesh et al [106] demonstrates the application of deep belief networks (DBNs) to describe intricate morphological characteristics of handwritten ancient Geez characters, achieving a recognition accuracy of 93.75%.

Arabic manuscripts containing hadith, tafsir, aqeedah, and the farthest Islamic sciences, which are essential in the development of Islamic culture. But to publish them, the manuscripts have to be typeset, which is not easy in this work. Alrehali et al [107] developed a method for scoring a text that appears in photographs of manuscripts or manuscripts and scoring them, using the scores for copying, so that it is easy to use them for further research. The essential processes of their algorithm are as follows: 1) Preprocessing (image enhancement); 2) Dividing the manuscript image into lines and words; 3) Making the dataset of Arabic characters; 4) Classification of texts.

In the classification step, they deploy a convolutional neural network (CNN) on three created datasets with an accuracy of 74.29% to 88.20%.

This study [108] goes ahead to investigate the various feature distances of Shui character image sample images so as to define and determine the least value of inter-class spacings and the optimal count of the clusters. With support and information from the classifier model, the ideal clusters in the clustering model are established, and a measure that combines information entropy adjustment and the parameter of convergence clustering unsupervised Shui character picture sample is estimated, and its effectiveness over its application is estimated

This paper [109] focuses on the segmentation and recognition of Devanagari characters using neural networks. It has been found that a hybrid feature extraction technique, which is geometric and statistical in nature, is employed. In the case of geometric feature extraction, directional features of the skeletonized character image are used. In contrast, the statistical feature technique depends on pixel density and Euclidean characteristics of a skeletonized character image. Classifiers such as Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) are used for the classification tasks. However, the SVM has greater accuracy than the Multilayer Perceptron.

This study [110] focuses on a transfer learning approach using deep convolutional neural networks for recognizing handwritten Devanagari alphabet symbols based on CNN-based feature extraction. This research uses AlexNet, DenseNet, Vgg, and Inception ConvNet as static feature extractors. They performed 15 epochs for each of the following models: AlexNet, DenseNet 121, DenseNet 201, VGG 11, VGG 16, VGG 19, and Inception V3. The results started with the epoch duration. Inception V3 surpassed the top accuracy level, achieving 99% on average, with a duration of 16.3 minutes per epoch, whereas AlexNet was the fastest, 2.2 minutes per epoch, but recorded 98% accuracy.

In the present work [111], new machine learning techniques are presented, allowing the automatic detection of the medieval signs (XI – XVIII centuries) that are placed upon the stone walls of St. Sophia Cathedral in Kyiv, Ukraine. A new computer-generated dataset that is related to carved Glagolitic and

Cyrillic characters (CGCL) was collected and prepared for recognition and prediction by making use of machine learning procedures. The dataset combines approximately 4,000 images which correspond to 34 different types of letters. The exploratory data analysis performed on the CGCL and the notMNIST datasets revealed that the recognition of the carved letters was difficult even when density discriminating the attention using t-SNE dimensionality reduction techniques owing to the fact that the representation of letters in stone carving is less effective than that done with a hand. Both Multinomial logistic regression (MLR) and two-dimensional convolutional neural network (CNN) models were employed. AUC values obtained under the ROC of the MLR model were not less than 0.92 and 0.60 for the notMNIST and CGCL datasets, respectively. It was found that the AUC value of the CNN model also reached about 0.99 for both the notMNIST and CGCL datasets.

The present work [112] attempts to describe new techniques in the recognition of Assyrian cuneiform characters by improving the limitations in the use of OCR to classify the symbols, thus presenting a new approach to the understanding of Assyrian letters by simplifying the problems of reading complex characters. The data set used contains sixteen patterns to encode every possible shape and orientation defined for the cuneiform signs, taking into account that every character is composed of several symbols. For the description of the image analysis in the identification problems, polygon approximation techniques are applied for the construction of feature vectors. The presented approach enables achieving classification ratios close to 91 % relying on the recognition algorithm of the feature vector used.

In this paper, Lamyaa Sadouk et al [113] suggest two methods using deep learning for the recognition of handwritten Tifinagh characters: Convolutional Neural Networks (CNNs) and Deep Belief Networks (DBNs). Both networks were trained and tested on the AMHCD database, a collection of handwritten characters. The deep belief networks (DBN) achieved an accuracy of 95.47% while the convolutional neural networks (CNN) exceeded other methods with an accuracy of 98.25%.

This study [114] aims to extend the elaboration tackled in the previous ones by creating a system for the character recognition of Brahmi, Grantha, and Vattezuthu from palm manuscripts of ancient Tamil documents, interpreting the text, and machine translating it into the present-day Tamil digital text format. While many scholars develop different algorithms and techniques for character recognition in other languages, making an equivalent for ancient scripts remains a significant challenge. The image recognition technology has come close to perfection in scanning texts in English and other languages. However, no Optical Character Recognition software is capable of reliably digitizing printed Tamil text. Very few people understand the ancient scripts, and even fewer are trying to write them out by hand. The proposed solution seeks to solve this problem by converting inscriptions and palm manuscripts into a digital text format in Tamil, thus rendering the entire corpus of Tamil historical texts available in the digital age. It replaces the OCR with Tamil Unicode representation. The technique consists of several stages: i) Image preprocessing, ii) Feature extraction, iii) Template Matching (Recognition), iv) Text to Image Destruction (in case there is a need), and also other operations necessary for accurate text conversion. The performance of this approach in terms of first pass recognition accuracy, for Brahmi script characters, was 91.57% with the use of neural networks and conventional image zoning techniques. The work on the second phase of the Vattezhuthu character set is yet to commence. The Vattezhuthu script has an accuracy of 89.75 % in conversion. Table 1 illustrates the studies of existing work in the literature, including the methodology adopted, benchmark datasets used, and study outcomes.

Table 1 Character recognition in research

No	Author(s)	Title/Publisher	Year	Data set	Method	Accuracy
1	Maheswari et al	An intelligent character segmentation system coupled with deep learning-based recognition for the digitization of ancient Tamil palm leaf manuscripts/ Springer Nature	2024	Agasthiyar Vaithiya Kaviyam, Ramayanam, Thruvilayadal, and other medical palm leaf sets	Modified CNN with 125 classes	96.04%

2	S. Saini et al	Construction of Handwritten Indus Signs Dataset Employing Social Approach/ Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII)	2024	Indus signs	Convolutional Neural Networks (CNN)	70%
3	S. Elshehaby, M et al	A Deep Learning Approach to Cuneiform Symbols Recognition/ Advances in Science and Engineering Technology International Conferences (ASET)	2024	Hammurabi cuneiform symbols	VGG16	99%
4	Y. Gong et al	Handwritten Chinese Character Recognition in Ancient Books Based on Improved YOLOv7/ 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)	2024	Guofeng chapter of the Book of Songs	YOLOv7	100%

5	Graham West et al	A deep learning pipeline for the palaeographical dating of ancient Greek papyrus fragments/ Association for Computational Linguistics	2024	Greek papyrus	ResNet	79%
6	C. Varshini et al	Tamil and English Handwritten Character Segmentation and Recognition Using Deep Learning/ International Conference on Communication, Computing and Internet of Things (IC3IoT)	2024	Mendeley website Tamil Documents	ResTANet	99.37%
7	R. Thendral et al	Enhancing Tamil Handwritten Character Recognition Using Multimodel/ Deep Learning 10th International Conference on Communication and Signal Processing (ICCSP)	2024	Unconstrained Tamil Handwritten Character Database (uTHCD)	RESNET50, VGG16, and LeNet50	Not specified
8	S. Devendiran et al	Handwritten Text Recognition using VGG19 and HOG	2024	Tamil Handwritten Character Database	VGG-19 and HOG Feature Descriptors	98.34% and 91.2%

		Feature Descriptors/ IEEE				
9	Sunitha S. Nair et al	International Journal of Electronics and Communication Engineering (IJECE)	2024	EMNIST (Extended MNIST)	CNN-RNN with Bi-GRU	96.72%
10	Supriya Mahadevkar et al	Enhancement of handwritten text recognition using AI-based hybrid approach/ MethodsX, Volume 12, 2024, 102654, ISSN 2215-0161	2024	IAM and RIMES datasets	Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) with a Connectionist Temporal Classification (CTC) decoder	98.50% and 98.80%
11	G. Singh et al	A Deep Learning-Based Pre-Trained VGG19 Model for Optical Character Recognition/ IEEE	2024	Textual information from images	VGG19	94.96%
12	Prof., Jyoti et al	Enhanced Handwritten Text Recognition through Bidirectional LSTM and CNN Fusion/	2024	IAM and RIMES	Convolutional Neural Networks (CNN) and bidirectional LSTM	88.2% and 89.5%

		International Journal for Research in Applied Science and Engineering Technology				
13	B. Yogalakshmi et al	Adaptive Convolutional Neural Network Based Handwritten Tamil Character Recognition System/ IEEE	2024	Handwritten Tamil characters	Adaptive Convolutional Neural Network (ACNN)	95.6%
14	Dr. Vicky Nair et al	Deep Learning and Hybrid Approaches For Script-Independent Handwritten Character Recognition: Advancing Generalization Across Diverse Languages/	2024	MNIST (Latin), EMNIST, and custom datasets for complex scripts like Arabic and Hindi	CNN Model Hybrid Models CNN + SVM CNN + Bi-LSTM and Transfer Learning Models VGG16 ResNet50	98.4% 99.28% 99.45% 99.15% 99.67%
15	Humayun et al	A novel methodology for offline English handwritten character recognition using ELBP-based sequential (CNN)/ Springer Nature	2024	EMNIST	Deep learning-based ELBP-CNN approach	87.31%
16	Dhanya Sudarsan and Deepa Sankar	An Ensemble Neural Network Model For Malayalam	2024	Malayalam Palm leaf manuscripts (PLMs)	Combining fine-tuned	96.40%

		Character Recognition From Palm Leaf Manuscripts/ACM			CNN and BiLSTM	
17	Jindal et al	An optimized CNN system to recognize handwritten characters in ancient documents in Grantha script/ Springer Nature	2023	Grantha script	Convolution Neural Network (CNN)	99.30%
18	Chaitali Chandankhede and Rajneeshkaur Sachdeo	Offline MODI script character recognition using deep learning techniques/ACM	2023	MODI script	RestNet50 and Inception V3	94.552% 93.923%
19	Snehal Gaikwad et al	Recognition of offline handwritten Devanagari characters using new mask-based approach, histogram of oriented gradients and AdaBoost/ACM	2023	Devanagari script (V2DMDCHAR and ISIDCHAR)	The histograms of oriented gradients (HOG) and AdaBoost ensemble boosting method	98.43% and 98.68%
20	Rakshit et al	Comparative study on the performance of the state-of-the-art CNN models for handwritten Bangla character	2023	Bangla handwritten character (Ekush)	CNN (DenseNet 201, EfficientNetB0, Inception-ResNetV2, Inception V3,	ResNet152V2 is a suitable CNN model for the recognition of Bangla

		recognition/ Springer Nature			MobileNet V2, NASNetMobile, ResNet50, ResNet152 V2, Vgg-19, Xception, and EkushNet)	character
21	Khaparde et al	Enhanced Nature-Inspired Algorithm-based Hybrid Deep Learning for Character Recognition in Sanskrit Language/ Springer Nature	2023	Sanskrit characters Dataset 1 (46 sets of handwritten characters) Dataset 2 (numerals, vowels, and consonants)	Support Vector Machine (SVM), Convolutional Neural Network (CNN), DeepLabv3, Long Short-Term Memory (LSTM), Hybrid Deep Learning (HDL) Without Optimization and Improved Butterfly Optimization Algorithm (I-BOA)	Dataset 1: 85.9%, 87.8%, 92.1%, 91.7%, 92.3% and 93.1% Dataset II: 85.1%, 86%, 92.1%, 91.4%, 92.2% and 93.1%
22	Vijaya Krishna Sonthi et al	An Intelligent Telugu Handwritten Character Recognition	2023	Handwritten Telugu characters	Multi-Objective MayFly Optimization with	98.74%

		Using Multi-Objective Mayfly Optimization with Deep Learning–Based DenseNet Model/ACM			Deep Learning (MOMFO-DL)	
23	S. Ezhilarasi et al	Recognition of Characters using PCE based Convolutional LSTM Networks from Palaeographic Writings/IEEE	2023	Historic Tamil palaeographic writings	CNN-LSTM	97.9%
24	R. Jagannathan et al	ATCRI-21 Neural Network Based Character Recognition with Image Denoising in Ancient Epigraphs/IEEE	2023	Ancient Tamil characters (Brihadeeshwar temple, Thanjavur)	Convolutional Neural Network (Ancient Tamil Character Recognition in Inscription – 21) ((ATCRI-21))	80%
25	Islam et al	Manuscripts Character Recognition Using Machine Learning and Deep Learning/ Modelling	2023	Beowulf manuscript's Old English characters	Convolutional Neural Network (CNN)	98.86%
26	H. Sajjad et al	Learning Structural Similarities from	2023	Handwriting on papyri	DenseNet-121 and omniglot-	62% and 69%

		Handwriting on Papyri - An Application to Scribe Characterization/IEEE			trained CNN	
27	Sujith Kumar S et al	AI Based Tamil Palm Leaf Character Recognition/ International Journal of Scientific Research in Engineering and Management (IJSREM)	2023	Tamil manuscripts written on palm leaves	Convolutional Neural Networks (CNNs)	89%
28	Zhenbang Wang	Ancient character recognition with deep learning techniques/ EWA	2023	Oracle-MNIST	CNN	97.009%
29	Edward C et al	DeepScribe: Localization and Classification of Elamite Cuneiform Signs Via Deep Learning/ ACM	2023	Elamite Cuneiform Signs (Persepolis Fortification Archive)	DeepScribe, a modular computer vision pipeline	80%
30	Jyoti Madake et al	MODI Script Recognition System Using Zernike Moments and DCT Features/IEEE	2023	MODI script	Decision Tree, Random Forest, and KNN	97.77%

31	P. Thuvarakan et al	The Application of Convolutional Neural Network in the Context of Tamil Handwritten Character Recognition/IEEE	2023	Tamil characters	CNN	95%
32	B. S. Babu et al	Temple Inscriptions Recognition and Transliteration in Devanagari Script/IEEE	2023	Vatteluttu script	Deep Learning	84.12%
33	Priyadharshini.R et al	Handwritten Characters to Electronic Readable Text Using Deep Learning/IEEE	2023	MNIST	Convolutional Neural Network (CNN)and Support Vector Machine (SVM)	96.5% and 94%
34	Xiaoyu Yuan et al	A New Database of Houma Alliance Book Ancient Handwritten Characters and Its Baseline Algorithm/ACM	2023	Houma Alliance Book ancient handwritten characters	DCF-LAR (Decision-level Classifier Fusion with Adaptive Recognition)	84.8%
35	N. Sasipriyaa et al	Recognizing Handwritten Offline Tamil Character using VAE-GAN & CNN/IEEE	2023	Handwritten Tamil letters	Variational Autoencoder (VAE) and Generative	86%

					Adversarial Network (GAN)-VAE-GAN with CNN	
36	Shai Gordin	Optical character recognition for ancient non-alphabetic scripts/ open access government	2023	Latin transliterations of Akkadian	Deep Learning	89%
37	Savitri Jadhav and Vandana Inamdar	Convolutional Neural Network and Histogram of Oriented Gradient Based Invariant Handwritten MODI Character Recognition/ACM	2022	MODI script	CNN-based transfer learning and a global feature extractor histogram of oriented gradient	Not specified
38	Raghunath Dey et al	Offline Odia handwritten character recognition with a focus on compound characters/ACM	2022	Odia character images	RNN and CNN	86.56%
39	Gebremichael et al	OCR System for the Recognition of Ethiopic Real-Life Documents/ Springer Nature	2022	Ethiopic script	Gabor Filter and Principal Component Analysis (PCA),	98.33%

					followed by a Genetic Algorithm (GA) based on supported vector machine classifier (SVM)	
40	Tabita L et al	Isolated Handwritten Character Recognition of Ancient Hebrew Manuscripts/ Society for Imaging Science and Technology	2022	Handwritten ancient Hebrew text	Convolutional Neural Networks (CNN) [LeNet-5, AlexNet, VGG16, and ResNet50]	96,34% 94,44% 91,79% and 89,39%
41	M. Pandey et al	Character Recognition of Vedic Sanskrit using Deep Learning Algorithms/IEEE	2022	Sanskrit language	4-fold Convolutional Neural Network (CNN)	89.975%
42	Joseph, S., George, J.P.	Offline Character Recognition of Handwritten MODI Script Using Wavelet Transform and Decision Tree Classifier/ Springer Nature	2022	MODI script	Daubechies, Haar, and Symlet wavelets Decision tree classifier	Not specified
43	Feichi Han	Handwritten character recognition	2022	English Handwritte	CNN	96%

		with stable and adaptable convolutional neural network/ SPIE		n Characters		
44	R. Gayathri and R. Babitha Lincy	Transfer learning based handwritten character recognition of tamil script using inception-V3 Model/ ACM	2022	Tamil handwritten character offline dataset (Hewlett-Packard Lab: hpl-tamil-iso-char-offline-1.0.)	Inception-v3	93.1%
45	R. Fermanian et al	A Deep Convolutional Neural Network for Character Recognition in Ancient Syriac Manuscripts/ IEEE	2022	Syriac handwritten characters	Deep character recognition network	97%
46	Dhivya S and Usha Devi G	TAMIZHİ: Historical Tamil-Brahmi Script Recognition Using CNN and MobileNet/ ACM	2021	Tamil-Brahmi inscriptions	MobileNet	Vowels (8 class) with 98.1% accuracy. Consonants (18 class) with 97.7% accuracy and consonants vowels (26 class)

						with 97.5% accuracy.
47	R. Rizk et al	A Hybrid Capsule Network-based Deep Learning Framework for Deciphering Ancient Scripts with Scarce Annotations: A Case Study on Phoenician Epigraphy/IEEE	2021	Phoenician alphabets	Hybrid capsule network-based deep learning framework	98.91%
48	S. P. Deore	Devanagari Handwritten Compound Character Recognition Using Various Machine Learning Algorithms/IEEE	2021	Devanagari characters	K-Nearest Neighbours (K-NN) and Support Vector Machine (SVM)	96.62%
49	Arwa Hamed Salih Hamdany et al	Translating cuneiform symbols using artificial neural network/ TELKOMNIKA Telecommunication, Computing, Electronics and Control	2021	Sumerian cuneiform symbols	Multi-Layer Perceptron (MLP) neural network	100%
50	P. D. Devi and V. Sathiyapriya	Brahmi Script Recognition System using Deep Learning	2021	Tamil Brahmi inscription		

		Techniques/ IEEE				
51	Bannigidad, P., Gudada, C.	Historical Kannada Handwritten Character Recognition Using Machine Learning Algorithm/ Springer Nature	2021	Historical Kannada handwritten scripts.	LDA, K-nearest neighbour (K-NN), and SVM classifiers	68.4% 85.7% and 87.5%.
52	Siranesh Getu et al	Ancient Ethiopic manuscripts character recognition using Deep Belief Networks/African Journal Online (AJOL)	2020	Ethiopic ancient manuscripts	Deep Belief Networks (DBNs)	93.75%
53	B. Alrehali et al	Historical Arabic Manuscripts Text Recognition Using Convolutional Neural Network/ IEEE	2020	Arabic manuscripts (Hadeethe, Tafseer and Akhidah)	Convolutional Neural Network (CNN)	Accuracy ranges between 74.29% to 88.20%.
54	Hongshuai zhao et al	Improvement of Ancient Shui Character Recognition Model Based on Convolutional Neural Network/ IEEE	2020	Shui characters in ancient book	Convolutional Neural Network (CNN)	70%

55	Khanderao, M.S., Ruikar, S	Character Segmentation and Recognition of Indian Devanagari Script/ Springer Nature	2020	Devanagari Characters	SVM (Support Vector Machine) and MLP (Multi Layer Perceptron)	99.8% and 97.4%
56	N. Aneja and S. Aneja	Transfer Learning using CNN for Handwritten Devanagari Character Recognition/ IEEE	2019	Handwritten Devanagari alphabets	AlexNet, DenseNet 121, DenseNet 201, Vgg 11, Vgg 16, Vgg 19, and Inception V3	Inception V3 performs 99% accuracy while AlexNet achieving 98%accuracy.
57	Nikita Gordienko et al	Open Source Dataset and Machine Learning Techniques for Automatic Recognition of Historical Graffiti/ACM	2018	Glagolitic and Cyrillic letters (CGCL) and notMNIST	Multinomial logistic regression (MLR) and a 2D Convolutional Neural Network (CNN)	99%
58	A. M. S. Rahma et al	Recognize assyrian cuneiform characters by virtual dataset/ IEEE	2017	Assyrian cuneiform characters	Polygon approximation techniques	91%
59	Lamyaa Sadouk et al	Handwritten tifinagh character recognition using deep learning	2017	Tifinagh-IRCAM	Convolutional Neural Networks (CNNs) and Deep Belief	98.25% and 95.47%

		architectures/ ACM			Networks (DBNs)	
60	E. K. Vellingiriraj et al	Information extraction and text mining of Ancient Vattezhuthu characters in historical documents using image zoning/	2016	Brahmi, Grantha and Vattezuthu characters from palm manuscripts of historical Tamil ancient documents	Neural network and image zoning method	Brahmi script rate 91.57% and Vattezhuthu is 89.75%.

2.4.3 CONCLUSION

An overview of previously published work in the area of offline handwritten character recognition for ancient scripts points out that many strategies and methods have been implemented across many historical and geographic scripts. The mentioned studies witness the most promising development in the area, as CNN, Vision Transformer, and other deep learning technologies are commonly used to boost the recognition accuracy of ancient characters. The outstanding achievements are worthy of note as well, as the methods have produced accuracy rates from 62% to 100% depending on a particular script and the dataset utilized.

Moreover, research directed towards the recognition of ancient scripts such as Tamil, Cuneiform, and Sanskrit has also contributed significantly to the understanding of the application of deep learning models that can efficiently recognize complicated, degraded, and altered forms of characters. For instance, recognition tasks of Tamil and ancient Tamil scripts like Vattezhuthu and Brahmi have delivered high recognition results, including CNN-based models, ResNet, and VGG models on various datasets such as palm-leaf manuscripts and historical inscriptions.

Progress has been promising, but some ancients with lesser or non-tagged data scripts still present hurdles to overcome. Some practices have been implemented, such as employing neural networks along with image zoning, but these also point out the need for more studies to improve performance optimization in customization in more difficult situations. Again, the context of the language or the ancient Vattezhuthu still remains one of a kind in terms of recognition, for it requires sophisticated models for its distinctiveness and complex metadata.

To conclude, the majority of the current approaches focus predominantly on recognition tasks and suffer from the lack of error-correcting algorithms, which could lead to improved model recognition accuracy, especially for ancient scripts such as Vattezhuthu, from more diverse metrics and more standardization for the datasets used. In further studies, it is worth seeking ways to enhance models' generalization on several ancient writing systems in addition to the generation of richer and tagged data collections.

2.5 VISION TRANSFORMER CLASSIFIER

2.5.1 INTRODUCTION

Since deep learning and transformers' structural approach emerged as new paradigms, the field of computer vision has undergone massive changes over the years. The transformers were first created for tasks related to natural language processing. Since then, they have been used in image-related tasks like semantic segmentation, picture classification, and object detection, which has completely changed the field of vision. Transforming ancient handwritten character categorization is one such task that has greatly benefited from these advances. Vision transformers (ViTs) and Swin transformers built into transfer models have made a big difference in performance across key metrics like accuracy and robustness, while also providing scalability and flexibility.

Vision Transformers (ViTs) employ self-attention approaches to capture global information and have performed very well in visual tasks that require far-reaching dependencies. On the other hand, ViTs have had limitations due to their reliance on large datasets and computational capacity, which has led to the development of new strategies to overcome these limitations and apply them to niche areas and smaller datasets. However, Swin Transformers have used a hierarchical structure while keeping the global feature sharing part of local feature extraction. They have also improved the processing by adding a window-based self-attention mechanism. All of these models have done better than older models that used convolutional neural networks (CNNs) on several datasets and in several different application areas; they have raised the bar.

The goal of the review is to discuss the new direction that transformer-based architectures are bringing to the field of computer vision and show how useful they could be, especially for preserving and studying historical and cultural heritage.

2.5.2 VISION TRANSFORMER BASED REVIEW

Recent advances in deep learning and transformer-based architectures have significantly impacted computer vision, particularly the classification of ancient handwritten characters.

The Vision Transformer (ViT) has emerged as a revolutionary approach in computer vision. It adapts the transformer architecture, originally developed for natural language processing, to image classification tasks. One of the primary challenges addressed in training Vision Transformers is their dependency on large datasets.

The authors in [115] presented a comprehensive framework to successfully train ViTs from scratch, even on smaller datasets, by employing several innovative techniques:

1. CPE (CoordConv Position Encoding): This method uses 1×1 CoordConv to improve positional encoding, enabling the model to better capture spatial relationships within an image.
2. SPT (Shifted Patch Tokenization): Shifted patches are employed to enhance the locality inductive bias in ViTs, ensuring the model captures local features more effectively.
3. LSA (Learnable Self-Attention): A novel self-attention mechanism that incorporates learnable temperature parameters and self-relation masking, reducing unnecessary attention and improving focus on relevant features.

When these techniques were applied to the Tiny-ImageNet dataset, the authors observed an average accuracy improvement of 3.66%, with a maximum increment of 5.7%. These findings underscore the potential of ViTs for smaller datasets, paving the way for their application in specialized tasks such as historical text classification.

Further, the work in [116] highlights the efficiency of Vision Transformers compared to conventional convolutional neural networks (CNNs). Unlike CNNs, ViTs require fewer resources and parameters during training. This is particularly advantageous for tasks involving constrained computational budgets or smaller datasets. The study also emphasizes the transferability of ViTs, demonstrating their ability to perform effectively on various image recognition benchmarks after preliminary training on large datasets. This characteristic makes ViTs an attractive option for domains where annotated data is limited but transfer learning can be utilized.

The authors in [117] focused on improving the explainability and performance of ViTs by addressing the understanding gap in patch interactions. They identified three critical areas of development:

- Quantification of attention mechanisms to better understand patch-level interactions.
- Optimization strategies to enhance the performance of attention-based models.
- Establishment of novel architectures that balance performance with interpretability.

This work underlines the importance of creating explainable artificial intelligence (XAI) systems within computer vision, especially when applying ViTs to complex datasets such as ancient scripts.

Swin Transformers have introduced a hierarchical approach to the transformer model, making them particularly suited for computer vision tasks. Their unique shifting window mechanism improves efficiency and performance in image classification and object detection.

The study in [118] demonstrates the versatility of Swin Transformers by applying a Speech Swin-Transformer for emotion recognition from speech spectrograms. This model aggregates multi-scale emotional features, illustrating the flexibility of Swin Transformers across domains beyond computer vision.

In [119], the Swin Transformer model was adapted for paediatric oral health detection. By enhancing data augmentation and preprocessing techniques, the study achieved an accuracy of 85.34%, highlighting the potential of Swin Transformers for medical applications.

Further experiments reported in [120] evaluated the performance of Swin Transformers on datasets such as CIFAR-10, CIFAR-100, and MNIST. These experiments revealed that Swin Transformers consistently outperform other vision transformers in accuracy and memory efficiency, showcasing their practicality for resource-constrained tasks. In [121] Swin Transformer-Base, similar to its predecessor, optimizes the original transformer model for computer vision tasks through a hierarchical structure and shifting windowing mechanism that enhance image classification and object detection performance while maintaining computational efficiency.

In [122], the Swin Transformer was integrated into an improved Faster R-CNN architecture for vegetable recognition. The system achieved higher accuracy and efficiency by replacing the backbone network with a Swin Transformer, demonstrating its adaptability to diverse image recognition tasks.

Yucheng et al [123] introduced the new hybrid approach, Transformer-Convolution Network (TCNet). The TCNet combines the strengths of CNN and a Transformer module that solves the problem of extracting and comprehending local features in relation to one another. As a result, long-range dependencies and contextual relationships are maintained even with standalone convolutional networks. Here, the CNN block in TCNet is designed to capture local features of the input data set. Such features become the basis of the further processing performed by the Transformer module, where self-attention mechanisms are employed to determine relationships among different local features. In this manner, the Transformer module draws attention to different parts of the input feature map and thereby enhances long-range dependency, which the traditional CNN usually misses. This organized approach ensures that the features remain fully comprehensible while getting more detailed as they pass through the network.

The combination of the CNN and the TCNet modules of the transformer shows that there is a benefit in merging local feature extraction and global context understanding. This type of strategy greatly aids the ability of the model to detect text of different sizes, shapes, and orientations in natural scenes. Moreover, the Tween module improves the CNN module performance by remedying the weaknesses of CNN in building relationships across local features. Consequently, TCNet can attain better text detection accuracy, even with a limited dataset during training.

Gu and Min [124] present a new model that utilizes the Swin Transformer architecture and is tailored specifically to classifying 2D shapes.

The architecture embraces a unique two-stream approach such that while one of the streams is focused on shape images in their raw form, the remaining is designated towards the shape in their all-scale representation. However, by combining the various forms of feature sets, the model improves its proficiency with shape recognition. Self-Attention patches have been embedded into high-dimensional spaces, while input images have been segmented into non-overlapping parts, ensuring optimum performance by the Self-Attention modules.

Zheng et al [125] present a new recognition model built upon the Swin-Transformer architecture with Context-Transformer (CoT) attention blocks. In this context, it is significant to note that this architecture tolerates multi-layered feature extraction and enhances overall computational efficiency. The Swin-Transformer character recognition model utilizes a window-based mechanism as its core component. This window mechanism aids in the learning of features over cross-windows by combining global and local. Under this mechanism, deep networks are better equipped to handle finer details characteristic of ancient Chinese characters. The model has replaced the multi-head self-attention (MSA) that most self-attention modules utilize with a shifted windows-based MSA module. The purpose of this change is to avoid computational intensity by reducing the number of tokens while maintaining the ability to extract useful relationships from the data.

The model's internal framework commences by taking input images and splitting them up into patches that are not overlapping, akin to the procedure followed in a Vision Transformer (ViT). The patches are treated as a kind of token; to be more specific, patch features are determined by the raw pixel RGB values, resulting in an efficient means of packaging for the subsequent stages. LayerNorm components are placed before both the MSA and MLP modules, while residual connections make the training process less sensitive and promote better fluidity of information. In addition, there is a gradual downward scalability in the architecture of the structure for constant input feature maps. At the same time, the receptive field is increased, which is an important criterion when it comes to distinguishing intricate and delicate characters.

The performance of this method is validated through experimental results wherein the model can achieve a top-one accuracy of 87.25 percent and a top-five accuracy of 95.81 percent on the more meticulously curated dataset containing ancient Chinese characters. This method, as proposed, shows the possibility of using new architectures based on the Transformer for solving complicated tasks related to character recognition.

The researcher has utilized a novel approach for recognizing handwritten Devanagari numerals through a vision-transformer-based learning model, and this approach has been integrated by Rajpal and Garg [126] through using the Shifted Window (SWIN) transformer. The presented framework is a major improvement over manual processes in resolving the computing and spatial issues that are present in deep learning models, as well as achieving maximum recognition accuracy.

The hierarchical scaling built within the proposed model supports feature extraction at various scales, which is critical due to the diverse shapes and sizes of handwritten numerals. SWIN transformer uses shifted window approach for self-attention which capture both local and global features of the model by working across multiple parameter space between different image patches. The amalgamated use of diagonal window and shifting enhances the model's performance on understanding the structure of complex numerals with variable customs even further.

The study made a thorough assessment of the SWIN transformer over deep convolutional neural network architecture models including VGG-16Net, ResNet-50 and DenseNet-121. These were considered popular among researchers to contain a strong feature extraction ability for the recognition of numerals. The performance of the SWIN DCNN on other models was highly competitive, achieving an impressive numeral recognition accuracy of 99.20%. This speaks for its applicability as well as the promise of the transformer that this recognition model possesses for handwritten numerals.

One of the largest outstanding challenges in computer vision is the interpretation of handwritten text, particularly for complex languages, such as Urdu. Ganai and Khursheed [127] described a novel method of utilizing the BERT model for encapsulating inline images into handwritten Urdu text. The vision transformers' attention mechanisms were coupled to text images as a convolutional feature embedding, which constructs word embeddings as well. In turn, this allows a deeper understanding of the spatial interrelations between ligatures and diacritics.

The authors pointed out the possibility of using vision transformers for ligature patch feature extraction, which was incorporated into the transformer encoder. The segmentation of input images into patches is important, as each patch is sent for examination by an already trained object classifier to find important areas related to its contents and edges. Then these areas are examined with the help of a pre-trained network, for example, a CNN or MLP.

This combination produces a series of vectors fed into the transformer model for training and recognition. Again, embeddings of this kind enable the model to grasp the spatial relationships, which are necessary to understand the complexity of handwritten Urdu script properly. The model is improved by pre-training with publicly available datasets of UNHD and NUST-UHWR.

Recently, Vision Transformers (ViTs) have been developed as a replacement for CNNs for a number of tasks related to vision. The Pyramid Vision Transformer v2 (PVTv2) is another variation of vit, which has shown great promise with text recognition models. Unlike conventional ViTs, which apply fixed positional embeddings, PVTv2 uses convolution operations and, as a result, is more appropriate for encoding images of text with different widths. This quality of PVTv2 makes it ideal for dragon flag angling text, as characters of a particular language often have a high variety in width and level of complexity in terms of structure.

It has been confirmed by Wu, Li, and Wang [128] that the PVTv2 backbone can be used for the recognition of Chinese text when PVTv2 is modified for the given task, while providing a reasonable argument for the Chinese language, as well as suggesting a few other modifications. First, they mention that 32× down-sampling is carried out during text recognition machine usage, and this makes it the case that crucial details are lost. After that, they mention that parameter adjustment is needed for amending the angles of a given pvtv2 model. By employing those models, the backbone is capable of accurately recognizing Chinese characters while being efficient enough. Additionally, the capabilities of the PVTv2 vision transformer are also capable of accommodating a variety of character structures and sizes.

The writers also combined the PVTv2 backbone with two novel modules, namely, the Glyph-Aware Decoder and the Contrastive Visual – Textual Learning module. The Glyph-Aware Decoder employs glyph features to improve the ability to recognize Chinese characters while the Contrastive Visual – Textual Learning module fuses textual and visual features in order to raise the recognition accuracy.

Almutairi [129] proposed a framework in which the components of a newspaper are detected and classified using the image and text processing capabilities of the transformers, which are rather sophisticated. In this framework, a ViT is employed for the purpose of classifying the pages of the images input to the system as containing either a newspaper or a magazine. The basic concept in this technique is that CNNs use the concept of patching, which allows the image to be split into a number of patches that contain embedded spatial metadata before being sent into several blocks of transformers for classification. This method demonstrates the capability of ViTs to outperform the traditional classifiers based on CNNs. The results point to the great promise that transformers, most especially ViTs models, have in such tasks as document analysis and classification. This framework developed does not only make use of the advantages offered by transformers but also offers a good substitute of CNNs for the methodology of newspaper elements detection and classification.

Bensouilah et al. [130] have extended the combined use of gated multi-layer perceptron networks (gMLP networks) in the context of character-based handwriting transcription to deep learning applications. The gMLP created by the Google Brain team simplifies transformer's self-attention layers to a great extent. This architecture includes Spatial Gating Units (SGUs) which effectively model spatial interactions within the feature maps.

The gMLP architecture consists of a sequence of identical blocks, with each block carrying out spatial and channel projection to enhance feature representation. In the context of the proposed pipeline, the gMLP sits in between the CNN encoder and the RNN decoder whereby it decodes feature maps from the CNN. This functions to relate different spatial locations in the image and feature channels which optimizes the input. This improves the transcription performance because it reduces redundancy and increases emphasis on certain areas of input.

The study suggests gMLP-based networks are promising tools when character-based handwritten text transcriptions are incorporated, since they improve feature extraction and dependency modelling. Because the gMLP promotes efficient input learning by addressing spatial and channel-level interactions, it provides a more robust option for self-attention networks than self-attention networks do.

The use of Vision Transformers (ViTs) to identify the letters of the Arabic Sign Language (ArSL) was explored by Alnabih and Maghari [131], illustrating the potential of transformer-based models in picture classification tasks. In their study, unsupervised sign language data from ArSL2018, comprising 54,049 images belonging to 32 distinct ArSL categories, was employed to further train the ViT model. The processed images of the sign language enabled the fine-tuned ViT model to achieve reliable identification of the ArSL letters.

A pretrained ViT model was employed, previously trained on a larger dataset (google/vit-large-patch16-224-in21k). This type of model has gained its weight from a vast dataset; hence, to facilitate performance, all intermediate features of the model were frozen, and only a few lower-level task features were learned. In order to get the best outcome, the learning rate and batch size were finely adjusted.

The ArSL2018 dataset was employed and the ViT model produced an accuracy of 99.3%, also outperforming ResNet50 and MobileNetV2. Furthermore, the empirical assessment that made use of images taken from webcams established that the model was effective, achieving 100% precision on the selected testing samples. These results indicate that the ViT model is superior to traditional convolutional neural networks because it is effective in tasks such as identifying ArSL letters.

A transformer-based architecture that incorporates HTR traits for intricate diplomatic historical manuscripts with little data available is suggested by Mayr et al. [132]. The architecture has both convolutional neural networks and transformer encoders and decoders integrated into its core. The ResNet18 model is the backbone of the architecture, but is heavily modified for this task as its features were useful for recognizing detailed historical manuscripts.

In particular, for the case of 4000 training samples, the model lowered the Character Error Rate (CER) to 9.39% for contracted diplomatic texts and 12.07 % for texts with contracted abbreviations. These findings highlight the model's resilience as well as its efficiency of recognizing the complicated structures and diverse languages utilized in diplomatic historical papers.

Momeni and BabaAli [133] undertake a thorough examination of the Transformer-based approaches in artificial intelligence models using Arabic language offline handwritten text arguing two distinct models: the Standard Sequence-to-Sequence Transformer with Cross-Attention and the Transducer Transformer models. The traditional recurrent models are replaced by the Transducer Transformer, which employs a non-recurrent structure that seeks to improve parallelization and overall computational efficiency. The Cross-Attention Transformer, on the other hand, employs cross-attention mechanisms to enhance the modelling and decoding of sequences.

The suggested models achieve unprecedented results on the KHATT dataset, a leading benchmark for Arabic handwriting recognition. An impressive feature of this work is the capture of superior performance without the need to use very complicated pre- or post-processing techniques. The models use pretrained Transformer models, employing DeiT (Data-efficient Image Transformer) for image input and asafaya-BERT for text output, guaranteeing a high degree of accuracy and generalization of the model.

Out of the two methods, the Cross Attention Transformer had a far lower 18.45% mistake rate as compared to the Transformer Transducer which had a 19.76% error rate which shows how effective the cross-attention mechanism is at accurately predicting the relationships in Arabic handwritten text.

Wang et al. [134] undertook a thorough investigation into the recognition of license plates by proposing the integration of Cross Stage Partial Dilated Convolution Network (CSPDC) with YOLOv5. The integrative model was put through a bilateral analysis, which was conducted in association with the Swin Transformer-based YOLOv5. The Swin Transformer, being a hierarchical structure, utilizes a combination of smoothly shifted windows to process images efficiently by harnessing both local and global image features. Furthermore, 99.3% of accuracy was achieved on CSPDC-YOLOv5, whereas the plate recognition system's speed and precision reported a lower accuracy than that.

According to Ma et al [135], the STEF model, a Swin Transformer Enhanced Feature Pyramid Fusion framework, has also been developed for the purpose of detecting Dongba manuscripts. The Swin Transformer has been integrated into the STEF model to support its main aim of detecting multi-faceted and complex features embedded in historical Dongba texts. The model incorporates both global semantics of transformers and CNN-based local features using the Swin Transformer.

Swin Transformer works by dividing images into independent non-overlapping patches, which makes it easier to zoom into specific local characteristics without losing sight of the general picture. A window-based multi-head self-attention mechanism, W-MSA, and a window-shifted multi-head self-attention mechanism, SW-MSA, are used to achieve both feature enhancement and reduction of the computational burden. This new technique using the Swin Transformer expands upon what the STEF model can do, enabling it to extract information from complex features as well as analyse text at different scales, which is pivotal while working with Dongba manuscripts.

From the results obtained during the study, it can be said that the model has an 88.88% accuracy, recall of 88.65% and an F-measure of 88.76%. These results support the Swin Transformer's ability to solve complicated tasks such as feature extraction and show its suitability for the ancient script cultural patrimony preservation related activities.

The reviewed literature highlights the transformative potential of Vision Transformers and Swin Transformers in computer vision. By addressing key challenges such as data dependency, explainability, and computational efficiency, these models have set new benchmarks for image classification tasks. The integration of their strengths into hybrid architectures like HybridViTSwinModel marks a significant step forward, enabling superior performance in specialized applications like historical text classification. Continued research and innovation in this field are likely to expand the scope of these models, fostering advancements across diverse areas of artificial intelligence and machine learning.

2.5.3 CONCLUSION

The performance of computer vision has significantly improved due to the evolution of deep learning and transformer architecture. Particularly, the challenges of classifying ancient handwritten texts have benefited tremendously. ViTs and Swin transformers have emerged as powerful models with the potential to efficiently integrate global contexts with local hierarchies. Their ability to outperform traditional CNNs with higher accuracy and efficiency coupled with scalability has made them pivotal in almost all applications ranging from historical text classification, medical screening, and a lot more.

Transformer models in hybrid and domain-oriented architectures have advanced explainability, transferability, and resource efficiency. These models have been able to increase recognition performance in historical and multilingual texts in addition to operating well under limited resources, showing their versatility and transformational capabilities.

3. RESEARCH METHODOLOGY

3.1 INTRODUCTION

Dissecting any manuscript, inscription, or historical object is a challenge, but one that is rewarding, for it requires the usage of advanced technologies to ensure that stones of cultural significance are preserved and analysed as well as interpreted. These delicate and often damaged sources require advanced machine-learning models as well as multidisciplinary techniques for data extraction. This study provides a structured approach to recognizing and classifying ancient Vattezhuthu characters, mainly copper plates, palm leaves, and stone inscriptions.

This chapter outlines the comprehensive methodological framework employed in this research, providing a systematic overview of the steps involved. Each of these methods will be elaborated in detail in subsequent sections, but the key components are summarized here to give an overview of the process and its significance.

3.2 DATASET CONSTRUCTION

The foundation of this research lies in constructing a high-quality dataset of ancient Malayalam script Vattezhuthu samples. The primary sources include copper plates, stone inscriptions, and palm leaf manuscripts, sourced from prominent repositories such as the Hill Palace Archaeological Museum, the Tripunithura Palace, the State Archives Department, and the University of Calicut's Tunjan Manuscript Repository. These artifacts were carefully collected and digitized to serve as the raw data for further analysis.

3.3 GRAY SCALE CONVERSION

To ease the computational burden in image processing and analysis, the colour images collected were changed into grayscale images. By such conversion, the recoloured images are changed into another of differing complexity, gray, allowing one to pay attention to recognition-related structural and textural characteristics alone.

3.4 SUPER RESOLUTION (SR)

Due to the age of the artifacts, they often suffer damage, leading to them being captured in low-resolution images. Information that is crucial is thereby lost. Applying the Super Resolution techniques, missing portions were recreated, and image quality was enhanced so that finer details could be put in for examination.

3.5 NOISE REDUCTION TECHNIQUES

The manipulation or weathering of relics may cause them to have imperfections or scratches; this noise often makes the interpretation of historical artifacts harder. Due to this problem, multiple noise reduction methods were utilized, including the removal of Gaussian noise, salt-and-pepper noise, and speckle noise. Such techniques ensured correctness in the pre-processed images.

3.6 BINARIZATION

Binarization is the process of converting grayscale images into black and white, enhancing contrast, and facilitating feature extraction. In this study, the effectiveness of three prominent binarization algorithms—Otsu, Niblack, and Sauvola—was evaluated. Additionally, an improved method was proposed called GATED (Gaussian Adaptive Thresholding with Edge-Enhanced Dilation), specifically targeting the binary recognition of ancient characters. The Structural Similarity Index Measure (SSIM), Mean Square Error (MSE), and Peak Signal-to-Noise Ratio (PSNR) were used to measure how well these methods worked. The results demonstrated that the proposed approach outperformed the traditional techniques in terms of effectiveness.

3.7 SEGMENTATION

Segmentation accuracy is crucial for the determination of the Region of Interest (ROI) for the pre-processed images. A new algorithm for defining an ROI was created, which includes automatic contour line tracing and bounding box creation. This guarantees a proper text extraction to allow concentrated analysis.

3.8 RESIZING NORMALIZATION

For consistency across the dataset, all marked regions of interest were resized to a resolution of 160x160 pixels. This made the dataset usable for machine learning models while improving overall efficiency.

3.9 DATA AUGMENTATION

The limited quantity and diversity of ancient writing samples make it necessary to come up with new methods for data augmentation in order to improve the dataset. A novel stroke-based augmentation method named SBA

(Stroke-Based Augmentation) was carried out, which is much better than the usual techniques of flipping, rotation, and zoom. This advanced technique improves the recognition of Vattezhuthu cursive handwriting by emulating complex alterations to the stroke patterns.

3.10 HYBRID VISION TRANSFORMER-SWIN MODEL CLASSIFICATION

As a result of the varying forms of letters, the classification of early writing systems raises unique challenges. This challenge was resolved by developing a new hybrid Vision Transformer-Swin model named HybridViTSwin. This model combines the global self-attention characteristics of Vision Transformers (ViTs) with the local hierarchical self-attention features of Swin Transformers, allowing for efficient encoding of both global and local information. This new hybrid structure greatly increases the accuracy and reliability of ancient character classification.

4. PHASES OF DATA COLLECTION

4.1 INTRODUCTION

It is possible to figure out a country's ideology and history by looking at its proclamations, reforms, land and tax policies, treatment systems, daily records, artifacts, and justice systems. Historians would be wrong to ignore any of these things when studying these different countries. All of these, even the most ancient manuscripts, can nowadays be found. One such source of information is Vattezhuthu, which is believed to be one of the oldest writing systems. The Vattezhuthu script is found mostly in the rock-cut caves, metal plates, and palm leaves throughout the Kerala region. The recovery and preservation of such manuscripts are quite essential. But it must be borne in mind that such scripts, if maintained poorly, will be damaged or destroyed and will not be able to be used in the future. Reading such scripts opens an entirely different world of language, culture, and history of a people. Only a handful of people can read and understand this script. The use of the scripts provides the basis for investigating writing as an evidence-based technique of doing history. Dealing with these components is crucial in reconstructing a region's, its people's social, cultural, and political being. To explore this issue, it is important first to decipher the language appropriate for that world.

It is very challenging to capture the Vattezhuthu handwriting on various surfaces. Image acquisition is affected by factors such as the camera used, illumination sources, and the image's angle of reflection.

4.2 STRATEGIES FOR DATASET CONSTRUCTION

The task aimed at developing a fully usable, authentic dataset for the Vattezhuthu script ran into a number of formidable challenges, which called for novel approaches. The first and foremost challenge was the unavailability of a dataset that is relevant to the scope of this study and is available in the public domain. The processing of an image in image processing commences with the image acquisition, which is the retrieval of an image from an external source. This step is important and makes use of several capturing devices, such as scanners, cameras, and cell phones, which greatly affect the quality of images captured. The features of the device used for capturing the pictures and their parameters are given in Table 2.

Samples were collected from historical sources such as copper plates and stone inscriptions obtained from the Hill Palace Archaeological Museum and the palace located in the Tripunithura neighbourhood of Kochi, Kerala, India. Additionally, palm leaf manuscripts were sourced from the State Archives Department, Thiruvananthapuram, Kerala, India, and the University of Calicut Tunjan Manuscript Repository.

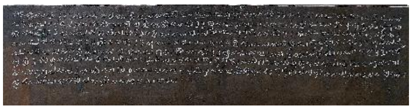
Table 2 Specifications of the image captured device

Property	Value
Horizontal Resolution	4032 pixels
Vertical Resolution	3024 pixels
Camera Model	I Phone 12
Flash Mode	No Flash
35mm Focal Length	26

4.3 GRAY SCALE CONVERSION

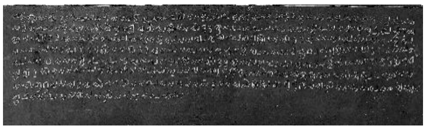
There are various advantages associated with converting Red, Green, and Blue (RGB) images to grayscale, which in turn enhances image processing tasks. Grayscale images have only a single intensity channel rather than the three channels (red, green, and blue) in RGB images; hence, the amount of data is highly simplified. The amount of data consumed is then reduced, leading to smaller storage space, reduced memory usage, and increased speed of computation, which is very helpful in the case of big volumes of data or time-sensitive cases. Grayscale images also improve the contrast by emphasizing the differences in intensity, which is useful in lightness and darkness operations, for example, edge detection, analysing textures, and thresholding. Overall, through loss of colour by converting the image into grayscale, this enables the elimination of most irrelevant colour information or variations that are noise contributed by the different colour channels, leading to improved extraction of features from the image.

In addition to that, it is easier for systems that do not make use of colour information to rely on grayscale images since they are likely to remain unchanged on different datasets. Many classical image processing and computer vision algorithms are optimized for grayscale inputs, further highlighting the practical advantages of this conversion. Other applications that can use grayscale images to their advantage include object detection, image segmentation, and OCR, since these tasks are based on how bright or dark certain areas are and not on their colours. Therefore, through grayscale conversion, such image processing workflows become more effective and efficient due to the reduction in noise and emphasis on the required structural features of the images. Figure 6 shows the original images and their corresponding gray-scaled images.



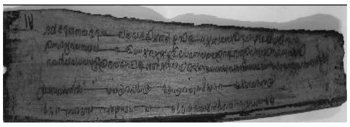
a) Copper plate

c) Palm leaf



d) Gray scaled image of (a)

e) Gray scaled image of (b)



f) Gray scaled image of (c)

Figure 6 Original images and its corresponding Gray scaled images

4.4 SUPER RESOLUTION (SR)

Super-resolution (SR) is a fundamental method in image processing that aims to recreate missing details from a lower resolution to produce a higher resolution image. Generally, it appears as though more pixels are given to the image to achieve the higher resolution that was initially brought down courtesy of down-sampling, blurring, and noise interference, the end results of which are destitute images. The idea behind super-resolving images involves increasing the size of the image by increasing the height and width of all the pixels in an image, which theoretically makes the picture clearer. The process not only enhances the clarity and detail of the image but also restores its lost information, which is particularly helpful in interpreting and analysing images from various fields, including medicine, satellites, and digital pictures.

5 NOISE REDUCTION TECHNIQUES

5.1 INTRODUCTION

Within the datasets for image recognition, the image normalization process has become a standard practice to improve the visual clarity by eliminating different types of interferences, such as speckle noise, Gaussian noise, salt and pepper noise, etc. This background shifts or blurs the original textural or structural information, thus complicating the correct understanding or interpretation of the thesis of the analysis. In relation to these problems, noise suppression problems are solved with the help of several state-of-the-art filters, namely, median filtering, Gaussian blurring, despeckling, etc., which are aimed at eliminating certain kinds of interference and enhancing the quality of the image. These techniques reduce noise damage to script images while retaining the important parts of the text or graphics.

5.2 REMOVAL OF SPECKLE NOISE

Speckle noise can be defined as a multiplicative noise that not only degrades the resolution but also the contrast of an image, leading to complications during numerous applications of image processing. This noise can be defined as:

$$I(i,j) = M(i,j) * N(i.j) \quad (1)$$

Where I(i,j) represents the resultant image, M(i,j) denotes the noise free image and N(i,j) refers to the speckle noise in an image.

Despeckle filtering techniques are applied in order to assist in minimizing the adverse effects that are produced by the speckle noise. These filters work in a manner in which they do not allow any noise to be added into the images while retaining the important buildings such as edges to define the structures within the image. An effective method to achieve this is the use of despeckling filters that replace each pixel value with the median value of a defined neighbourhood about the pixel. Such median filtering approaches manage to reduce noise without greatly causing edge blurriness.

5.3 REMOVAL OF SALT AND PEPPER NOISE

Often referred to as salt and pepper noise, impulse noise can be described as possessing maximum and minimum intensity measures. To eliminate or reduce such noise requires the use of good filtering, and in this case, emphasis will be placed on spatial domain non-linear smoothing techniques. Of these methods, the median filter is among the best in the removal of salt and pepper noise, fine line and edge detail, and general image sharpness level with minimum degradation of the critical image features.

The output of a median filter is obtained by first constructing a kernel window of every pixel with intensity values of image sub-elements, in this case, a kernel. For example, if P = 3 and the kernel measures 3×3 (9 pixels), then the median is taken of all the pixel values in the kernel area prior to filtering of the centre pixel. It is hypothesized that reducing the kernel size to 3×3 saves on computational time and effort, alongside being able to fight noise artifacts in the image. This logic can be described mathematically as follows:

$$D(m,n)=\text{median}\{Blur(S(p,q))\}, \text{where } (p,q) \in Kpq \qquad (2)$$

In the original image S, the notation S(p,q) denotes the value of the intensity of the pixel with coordinates (p,q). So, after performing the median blur operation, the value D(m,n) is expected to be in the denoised image for the specific pixel located at (m,n). All the pixel values within the kernel Kpq are collected. By sampling and integrating these pixel values, the median filter is able to mask out salt and pepper noise artifacts without affecting the context and detail of the image.

5.4 REMOVAL OF GAUSSIAN NOISE

Gaussian noise is a form of statistical interference whose Probability Density Function (PDF) is normally distributed. It should be noted that Gaussian noise is common to a number of imaging systems, where it is mostly introduced during the image capture stage as a result of sensor errors, electric interference, or environmental noise. In mathematical terms, this noise can be expressed as:

$$P(G) = \sqrt{1/2\pi\sigma^2}\ e^{-(g-\mu)2/2\sigma 2} \qquad (3)$$

where g is the pixel value, σ is the standard deviation, and μ is the mean.

The smoothing process is therefore able to eliminate or at least suppress high-frequency noise components, while retaining low-frequency components critical to the image, such as edges and textures. The smoothing operator has interesting features that are capitalized on to improve the quality of images.

6 BINARIZATION

6.1 INTRODUCTION

One of the most difficult issues with binarizing damaged historical manuscripts is selecting the best binarization threshold after implementing picture-enhancing procedures. The importance of binarization in the analysis of document images, especially of ancient manuscripts, cannot be overstated. It is a process that enables the separation of characters situated in the foreground and those located in the background by converting the images into two levels. Broadly speaking, there are two main classification schemes for the methods of binarization: global and local, or adaptive ones. Global binarization techniques rely upon a one-fits-all approach that stipulates one threshold to be applied to an entire image, but such methods do not perform well against historical manuscripts because the degradation is not always uniform, the background textures are not always the same, and the lighting is not always consistent.

In this section, the method of binarization proposed is evaluated against three known methods, namely Otsu [24], Niblack [25], and Sauvola, J. and Pietaksinen [26], focusing on ancient characters inscribed on copper plates, palm leaves, and stone surfaces. The techniques in use have been widely used in document image processing due to their effectiveness in various factors affecting documents with either faded images or with uneven light distribution over the surface of the document.

The performance of these methods in conjunction with the proposed strategy is measured using three widely used indicators: Structural Similarity Index Measure (SSIM), Mean Square Error (MSE), and Peak Signal to Noise Ratio (PSNR). These indicators play an important role in the assessment of the quality of images in terms of the parameters obtained from the images, like the degree of noise, clarity, and structure of the images produced after binarization.

6.2 LOCAL OTSU

Local Otsu binarization is a processed oriented variation of the (global) Otsu's method which entails localizing and adapting strategies for thresholding in image processing. Where the global thresholding approach chooses only one threshold for the whole image. Local Otsu divides the image into smaller partitions or windows that do not overlap.

An appropriate threshold is computed for each of those individual locations for which local Otsu windows have been drawn.

This adaptive approach makes it easy to make more reasonable and finer contextual thresholding decisions. This applies to cases when the illumination, contrast, texture, or other features of some parts of an image are quite different. Local Otsu determines the thresholds for a specified area, making it easier to control the zone of interest during the binarization process.

Local Otsu binarization helps in the segmentation as well as the extraction of objects or features contained in highly complex backgrounds. It is useful particularly in cases where there are variations in lighting across the image. Further, photographs with textures and contrast or areas with a change in density make good use of this methodology. The self-adaptive features of the Local Otsu algorithm also increase the efficiency of detection and outlining of certain objects or features in the image. It improves details, curves, and complicated structure extraction by modifying the thresholding technique according to the characteristics of the given neighbourhood area. This method improves the quality of object segmentation and background removal operations quite significantly.

6.3 NIBLACK

The Niblack algorithm computes a threshold optimal for a pixel by taking into account the average value and standard deviation of the pixel intensities of the small area which is also referred to as the neighbourhood of the pixel. If the intensity of this pixel exceeds this threshold, the pixel is classified as foreground. Otherwise, it is classified as background.

Niblack thresholding is very advantageous to images where there is non-uniform illumination or varying background intensity. It helps in handling changes in contrast and light.

6.4 SAUVOLA

The Sauvola algorithm is a local thresholding algorithm that is automatic and dynamic, while image binarization is done. Its operation includes estimating the thresholds of every pixel in the picture by taking into account its adjoining pixels, which are present in the defined area in the image. The Sauvola algorithm is based on the information that means and standard deviations of the values of pixels, which lie within a small prespecified sliding window or neighbourhood surrounding every pixel. This window moves across the image, and the pixel data around the imaging is analysed. The mean value stands for the mean intensity level, and the standard deviation is the mean value of the degree of dispersion or scatter of the pixel's intensities within that area.

The statistical factors are used for the calculation of the binarization threshold using the Sauvola formula:

$$\text{Threshold} = \text{Mean} * (1 + k * ((\text{Standard Deviation} / R) - 1)) \quad (4)$$

Here, 'k' represents a parameter known as the Sauvola coefficient, which acts as a sensitivity control parameter for the binarization process. It adjusts the threshold value computed from the mean and standard deviation, affecting the algorithm's adaptability to different image conditions and textures.

The 'R' term in the formula denotes the maximum pixel intensity range within the image. The k-value essentially governs the sensitivity of the binarization process. A higher k-value results in a more adaptive threshold that is sensitive to smaller local variations in pixel intensity, potentially capturing finer details but also increasing the likelihood of including noise or artifacts in the binary output. Conversely, a lower k-value produces a less sensitive threshold, which might overlook subtle variations but provides a smoother binary output with reduced noise sensitivity.

6.5 PROPOSED METHOD- GAUSSIAN ADAPTIVE THRESHOLDING WITH EDGE-ENHANCED DILATION (GATED)

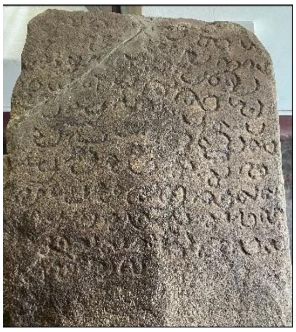
The proposed method, named GATED, includes systematic iterations targeting the enhancement of the quality of the images of text taken in the first place. The grayscale transformation is an important first step for the other enhancement processes that follow. Next, the super-resolution method is applied to boost the resolution of the image. Inherently, super-resolution techniques employ one or more low-resolution images to reconstruct a higher resolution image in order to enhance image quality. In short, this process includes estimating the pixels in order to boost the amount of detail and clarity of text in the images that were taken. After the resolution enhancement stage, a variety of techniques aimed at polishing the image are also applied. To smooth the images, Gaussian smoothing, median, and despeckle filtering are used.

After these necessary preprocessing stages, the process of the adaptive thresholding technique begins. This method computes the threshold value as the sum of the values of the nearby pixels multiplied by a weight factor, which is determined using a Gaussian window. This adaptive thresholding technique focuses on the local variations of the image intensities, and hence it becomes more flexible and sophisticated for binarization of images.

To continue with the method, the edges in the horizontal (X-direction) and vertical (Y-direction) directions of an image are determined using the Prewitt operator, which is an edge detection filter. This step is very important for the purpose of locating and separating the borders of any textual components in order to locate and separate the borders of the textual parts.

Finally, the text structural elements and appearance are enhanced using a dilatation operation. Dilation emphasizes the text areas and makes enhancement of the readability and clearness of the text areas relative to the original ones by expanding the shapes in the picture. After the last image processing round, this dilatation technique also helps to improve and refine the visual appeal of the text.

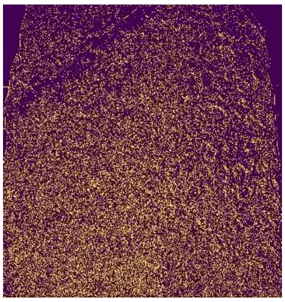
Figures 7,8, and 9 illustrate the binarized stone inceptions, copper plate and palm leaf using the above methods.



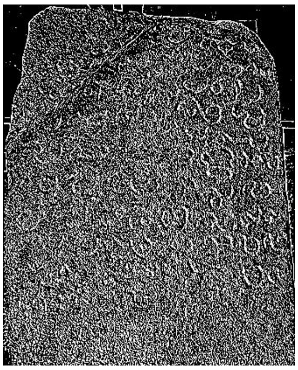
a) Original Image

b) Local Otsu

c) Niblack



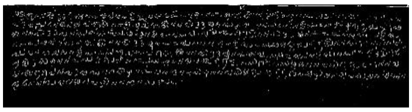
d) Sauvola



e) Proposed method

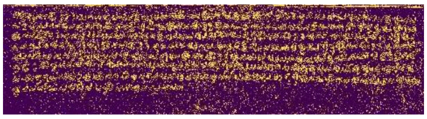
Figure 7 Binarized stone inceptions of the above methods

a) Original Image

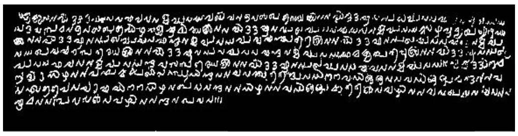


b) Local Otsu

c) Niblack



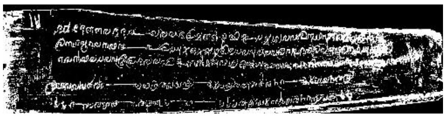
d) Sauvola



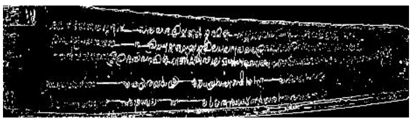
e) Proposed method

Figure 8 Binarized copper plate of the above methods

a) Original Image



b) Local Otsu



c) Niblack

d) Sauvola

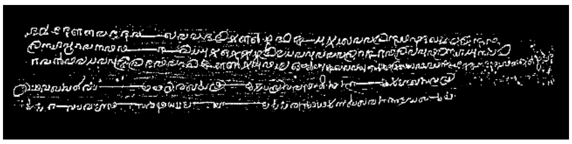


Figure 9 Binarized palm leaf of the above methods

6.6 EXPERIMENTAL RESULT AND PERFORMANCE EVALUATION

The achieved outcomes are assessed through three metrics, which include mean squared error (MSE), structural similarity index measure (SSIM), and peak signal to noise ratio (PSNR) of binarized images of copper plate, stone, and palm leaf.

On the contrary, when the value of PSNR is low, this implies that the enhancement quality is poor, and when the peak signal-to-noise ratio (PSNR) is high, this implies that the enhancement quality is good. As a result, it can be said that, in this case, the mean squared error, or MSE, estimates dissimilarities between two images that are similar in some aspects, for example, an original image and an enhanced one.

This is, the mean squared error tends to estimate the difference in image quality, the greater the MSE, the lower the quality enhancement and so, the lesser the value of MSE, the greater the enhancement of the quality.

$$PSNR = 10.\log_{10} \left(\frac{MAX^2}{MSE}\right) \quad (5)$$

Where:

- MAX: The maximum possible pixel value of the image. For an 8-bit image, this is usually 255.
- MSE: The Mean Squared Error between the original image (I) and the reconstructed or processed image (K), defined as:

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[I(i,j) - K(i,j)]^2 \quad (6)$$

where I(i,j) is the pixel value of the original image and K(i,j) is the pixel value of the reconstructed image.

SSIM [27] functions as a metric for evaluating structural information, including brightness, contrast, and structure. It is officially defined as:

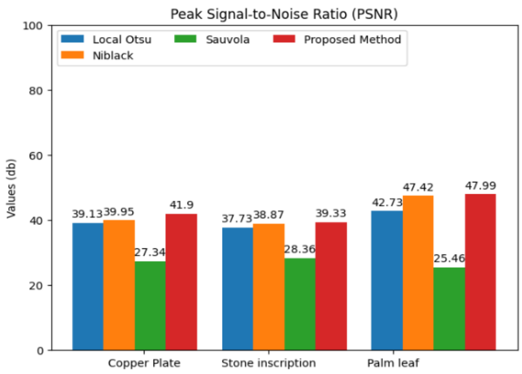
$$SSIM(x,y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (7)$$

Where μ_x is the pixel sample mean of image x, μ_y is the pixel sample mean of image y, σ^2_x is the variance of image x, σ^2_y is the variance of image y, σ_{xy} is the covariance between images x and y and C1, C2 is the stabilization constants to prevent division by zero, typically small positive values. In table 3 compared result is explained.

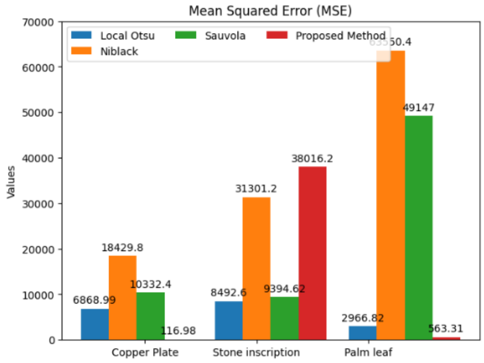
Figure 10 shows PSNR, MSE and SSIM values of the copper plate, stone and palm leaf images binarized by Otsu, Niblack, Sauvola and proposed method.

Table 3 PSNR, MSE and SSIM values of the copper plate, stone and palm leaf images binarized by various techniques

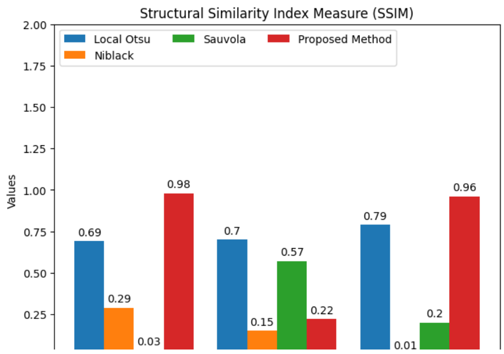
Material	Binarization methods	Performance Evaluation metrics		
		PSNR	MSE	SSIM
Copper Plate	Local Otsu	39.13	6868.99	0.69
	Niblack	39.95	18429.76	0.29
	Sauvola	27.34	10332.44	0.03
	Proposed Method	**41.9**	**116.98**	**0.98**
Stone inscription	Local Otsu	37.73	8492.6	0.7
	Niblack	38.87	31301.17	0.15
	Sauvola	28.36	9394.62	0.57
	Proposed Method	**39.33**	**38016.19**	**0.22**
Palm leaf	Local Otsu	45.38	5301.38	0.86
	Niblack	47.42	63550.44	0.01
	Sauvola	25.46	49146.98	0.2
	Proposed Method	**47.99**	**563.31**	**0.96**



a) PSNR comparison of local Otsu, Niblack, Sauvola and proposed method on copper plate, stone inscription and palm leaf.



b) MSE comparison of local Otsu, Niblack, Sauvola and proposed method on copper plate, stone inscription and palm leaf.



c) SSIM comparison of local Otsu, Niblack, and proposed method on copper plate, stone inscription and palm leaf.

Figure 10 PSNR, MSE and SSIM values of the copper plate, stone and palm leaf images binarized by Otsu, Niblack, Sauvola and proposed method

6.7 CONCLUSION

Binarization of ancient manuscripts has proven to be challenging and complex due to many noise types and degradation as well as due to the different physical and chemical properties of the materials used for preservation. This study focused on three different carrier forms, namely stone, palm leaves, and copper plates, which indeed posed different problems. For instance, stone inscriptions are always weather-beaten and have an uneven surface, palm fronds become brittle and change colour easily, while copper plates become discoloured and corroded. To address this variety of problems, advanced noise reduction technology and critically reviewing the more well-known binarization methods were required so that the new methods could be developed.

The adapted techniques outperformed available frameworks in enhancing the quality of binarized content across all three identified media, which was noted during different experiments and analyses using metrics such as MSE, SSIM, and PSNR. The results depicted the success of this approach in reducing noise and improving text clarity while also demonstrating its ability to ensure the reading and accuracy of the historical manuscripts.

This research provides a dependable and fast method for binarizing ancient manuscripts, greatly enhancing document image analysis and cultural heritage preservation. This study improves the effectiveness and accuracy of the binarization processes, thus claiming greater efficiency in the digital management of textual heritage. Therefore, it guarantees that such artifacts will not be lost but used and preserved for the future.

7 SEGMENTATION AND NORMALIZATION

7.1 INTRODUCTION

Image segmentation is a deeply understood stage in computer vision and image analysis. Segmentation is the basis of pattern recognition, medical imaging, and object detection. When segmenting an image, contour-based segmentation techniques are of great significance due to their efficiency and precision in targeting regions of interest (ROIs) within it. These methods enable the accurate extraction of objects from

exceedingly complicated visual environments by taking into account geometrical and spatial properties such as contours.

Image enhancement, de-noising, and other methods aimed at improving contours to be used in contour detection are known as pre-processing techniques and are the first step in the segmentation process. Once the contours have been pre-processed, they can be obtained as smooth curves or a series of dots that define the edges of the parts. Especially the contour detection methods employed here, and turning the image into a series of contours, are robust in their use, with each contour representing an ROI.

7.2 PROPOSED METHOD

The method of using specific contours to find the bounding box of the contours is a reliable method that is widely used when working with real-world segmentation tasks. It, in a way, simplifies the contour representation, as well as precisely determines the area of interest. The bounding box, being an adaptable tool, provides means for further uses such as measurements, classifications, and tracking of various objects. Also, it is well-suited for image segmentation tasks since it is able to integrate well with various image types and retain the content structure of the scene. It streams out the workload but still retains the boundaries of the salient features of the segmented areas by robustness to contour geometry. This method is able to show the increase in performance as well as the accuracy of the segmentation tasks in a total of applications.

The proposed technique revolves around bounding box construction and contour analysis to divide previously analysed images into regions of interest. Elucidated below are the specific steps that form the contouring process:

Locating Contours

Once the given image has been input into the system, the application will produce contours, which are several points that show the closure of a curve around areas of similar colour or intensity.

Contour Selection

A close contour that is relevant to particular identified contours is retrieved in order to enhance the final results. In order to ensure that only significant contours suitable for further operation are retained, parameters such as contour shape, dimension, and extent are set.

Formulating bounding boxes

For the purpose of image recognition, the regions of a contour which have been obtained are targeted while being encapsulated by the rectangles. Such conditioning includes using the smallest rectangle or bounding box so that the curvature is surrounded entirely along with specifying the coordinates for the recognition region.

Segmentation by Region of Interest (ROI)

ROI segmentation is performed using the bounding rectangles that were created with contours as masks. They act as boundaries for such downstream tasks like feature extraction and object recognition. Results of segmented characters samples are shown in figure 11.

Figure 11 Results of characters segmented from the sample input image

Validation Process- Visual Inspection

When the ground truth data is not available or when the segmentation quality needs to be judged by human standards under the guidance of a curator, the extracted ROIs are manually validated. Each ROI is examined by the curator to determine its correctness and applicability to the segmentation goals.

7.3 CONCLUSION

The developed ROI methodology integrates automated contour tracing, bounding box generation, and manual checking, resulting in the effective segmentation of ROIs in the pre-processed images. In this method, contours are carefully extracted using a step-by-step approach that employs a number of pre-processing processes followed by the bounding box estimations for the required contour. The manipulation of image data consists of auditing by the curator, ensuring the quality control is better than the automated ways, and ensuring the suitability of the extracted ROIs for the analysis.

7.4 RESIZING NORMALIZATION

After isolating the regions of interest in the input images, all of the images were resized to 160x160 pixels to create consistency in the whole dataset. The decision to select such a resolution is driven by the desire to balance feature preservation with efficiency. Such a resolution dramatically enhances the efficiency of further processing while maintaining the shape, structure, and outline of character images. The standard dimensions of 160x160 pixels have been shown to be efficient not only because they assist in the preparation of a dataset, but they also help in addressing the concern of overfitting in deep learning models and improving computational efficiency.

8. DATA AUGMENTATION

8.1 INTRODUCTION

Large and diverse datasets are particularly important for building models in machine learning, specifically for optical character recognition (OCR) tasks. For modern scripts, the creation of effective recognition systems is easily ensured due to the availability of large digitized corpora. However, it is not the case for several ancient scripts where the shortage of information is a large obstacle. The lack of ancient manuscripts and decoration creates challenges in gathering enough information, if not impossible, and therefore prevents the development of the OCR for such scripts. This challenge can be addressed through data augmentation, which allows for the resilient change of existing datasets to create new and varied ones. By addressing overfitting and promoting robust statistics, data augmentation streamlines the training process and boosts model performance tremendously.

Data augmentation tackles the problem of overfitting by encouraging the design of more robust models [39]. The majority of augmentation techniques include, but are not limited to, warping, rotation, flipping, cropping, shifting, smoothing, colour dilation, and adding noise. While these models hold true to modern scripts, they run the risk of becoming butchered to handwritten forms and letter forms of ancient scripts, where every detail would matter and be important. Such transformations become a risk of misrepresentation and will compromise the quality of the recognition models.

Here, the method approaches the augmentation of the characters' handwritten portions in a stroke-based fashion in which both structure and semantics are preserved. In this approach, data is enhanced without violating the geometric properties by concentrating on specific features of the character's strokes, thereby ensuring that enhanced data does not distort the original shapes. This approach widens the variety of datasets but, at the same time, maintains the purity of the ancient script and its legibility, making it suitable for teaching OCR algorithms based on a small amount of historical data.

This methodology is built on two main techniques. The first technique is concentrating on the random selection of regions within handwritten characters, followed by regulated expansion and contraction of the stroke's thickness. This random variation mimics the usual free-style handwriting variety, thus enlarging the training dataset and allowing the models to better adapt to numerous different handwriting styles. The second technique describes the stroke orientation. This approach normalizes the creation of characters by changing stroke direction and thickness and modifying the thickness of each stroke to accommodate changes in direction while allowing the samples to remain as close to the original writing as possible.

Utilizing the stroke-based augmentation of the handwritten portions as the baseline, this method utilizes geometric measures to determine the success of the techniques in preserving the original form and geometry of the characters. This ensures that the changes maintain the integrity of the document, resulting in enhanced, richer, and more accurate data for OCR training models.

8.2 PROPOSED METHOD- STROKE BASED AUGMENTATION (SBA)

As there is no established database of old handwritten Vattezhuthu script characters, it has to be developed into a moderately sized dataset by sourcing the information from different places. These includes sources such as copper plates, palm leaves, stone inscriptions and so on.

There are many ways to apply data augmentation, which relies on the strokes of the handwritten characters.

These augmentation techniques are split into two broad categories. With the first approach, areas of a handwritten character are randomly selected, and then alterations to the thickness and thinness of the stroke are introduced.

The second strategy is the stroke writing direction in a letter. The proposed method identifies the stroke directions and then modifies them by varying the thickness and thinness accordingly.

8.2.1 LOCALIZED THICKENING AND THINNING OF STROKES

In order to achieve the necessary regional thickening and thinning of the strokes, random image sections were constructed. These sections were defined as square areas with random positions of their centres and some prescribed width. The degree of change that was necessary was affecting both the quantity of parts and its maximum size.

The procedure to achieve the thinning of the stroke was performed through repeated morphological erosion but this was done only in those areas that were set out to be random. Erosion was performed with a cross-shaped structural member to enable thinning out but retaining the connection of the strokes.

Consider I as the binary image and B as the structural element which in particular is a cross shaped kernel measuring 3x3. The coordinates of the pixel (x, y) give the location of the individual in an image.

The following is a definition of how the structural element B erodes a binary image I at a pixel (x, y):

$$I_{eroded}\,(x,y)=\wedge_{(i,j)\in B}\, I(x+i,y+j) \qquad (8)$$

where: ∧ represents the logical AND operation. (Equivalent to the minimum in binary form, since 0 corresponds to background and 1 to foreground).

(x+i, y+j) represents the pixel positions within the neighbourhood of (x, y) defined by the structuring element B.

The picture pixel (x, y) is where the starting point of the structuring element B is.
As you go through each part (i, j) of the organizing element B, make sure that all of the pixels in the image I that go with it are in the foreground (i.e., 1 in a binary image).
The pixel (x,y) in the picture I eroded stays in the foreground (i.e., 1) if all of these pixels are in the foreground. If not, it goes to background (i.e., 0).

Figure 12 shows the results of the binarized input Vattezhuthu character ‘pa’.

a)

b)

Figure 12 a) Binarized input image ‘pa’. b) The thinned regions were highlighted in different colour for clear differentiation

The stroke thickening procedure consisted of repeatedly morphologically dilating selected random locations in the image. The number of times the dilation was performed determined how thick the final image would be. Dilation expands the areas of the foreground in a binary image by padding the perimeters of the objects with more pixels. The size and shape of the structuring element define the degree of the expansion.

Dilation of a binary image I with the use of a structural element B at pixel coordinate (x, y) is expressed as:

$$I_{dilated}\,(x,y) = \vee_{(i,j)\in B}\, I(x-i, y-j) \quad (9)$$

Here ∨ represents the logical OR operation (equivalent to the maximum in binary form, where 1 corresponds to foreground and 0 to background).

(x−i, y−j) represents the neighbourhood around the pixel (x, y) based on the structuring element B. The input binarized character 'pa' and the resulting images are shown in figure 13.

Figure 13 Shows the thickening of the stroke at different regions of the input character pa

The following pseudocode clearly describes the steps of the proposed technique.

Input: Binary image I (handwritten image), num_portions, max_portion_size, process_type (either 'thinning' or 'thickening')

Output: Modified binary image with localized thickening or thinning

Step 1: Generate random portions of the image

function generate_random_portions(I, num_portions, max_portion_size):

```
random_portions = []
for i = 1 to num_portions:
# Randomly select a center point (cx, cy) within the image dimensions
cx, cy = random_point_within_image(I)
# Randomly select the size of the square region (portion)
size = random_value(1, max_portion_size)
# Define the square region around the centre point
x_start = max(0, cx - size // 2)
y_start = max(0, cy - size // 2)
```

```
x_end = min(image_width(I), cx + size // 2)
y_end = min(image_height(I), cy + size // 2)
# Save the region as a portion
portion = (x_start, y_start, x_end, y_end)
random_portions.append(portion)
return random_portions
```

```
# Step 2: Perform localized stroke thinning or thickening

function modify_strokes(I, random_portions, process_type):

    # Create a cross-shaped structuring element (3×3 kernel)
    B = create_cross_structuring_element(3, 3)
    # Iterate over each portion
    for portion in random_portions:

    x_start, y_start, x_end, y_end = portion

# Extract the region from the image
    sub_image = I [y_start:y_end, x_start:x_end]
# Perform thinning (morphological erosion) or thickening (morphological dilation)
    if process_type == 'thinning':
    modified_sub_image=iterative_erosion(sub_image, B)
    elif process_type == 'thickening':
    modified_sub_image = iterative_dilation(sub_image, B)
    # Replace the portion in the original image with the modified sub-image
    I [y_start:y_end, x_start:x_end] = modified_sub_image
    return I
# Step 3: Morphological operations (Erosion and Dilation)
    function iterative_erosion(sub_image, B):
    # Apply erosion iteratively to thin the strokes
    for iteration in range(num_iterations):
    sub_image = morphological_erosion(sub_image, B)
    return sub_image
function iterative_dilation(sub_image, B):
# Apply dilation iteratively to thicken the strokes
    for iteration in range(num_iterations):
    sub_image = morphological_dilation(sub_image, B)
```

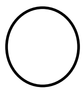
```
        return sub_image
# Main function to control the process
function main(I, num_portions, max_portion_size, process_type):
# Step 1: Generate random portions
        random_portions = generate_random_portions(I, num_portions, max_portion_size)
# Step 2: Modify the strokes based on the process type (thinning or thickening)
        modified_image = modify_strokes(I, random_portions, process_type)
        return modified_image
```

8.2.2 THICKENING AND THINNING BASED ON STROKE DIRECTION

The writing direction of the stroke of Vattezhuthu characters can be grouped into the following categories: Table 4 shows the direction of stroke and the corresponding character set.

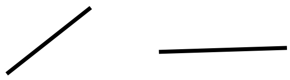
a) Clockwise curve

b) Closed round



c) Humps

d) Diagonal and horizontal lines



e) Anti-clockwise curve

Table 4 Direction of stroke and corresponding characters set

No	Stroke	Direction	Character
1	Clockwise curve	Top to bottom	Set-1
2	Closed round	Left to right	Set-2
3	Humps	Bottom to top	Set-3

4	Diagonal & horizontal	Left to right	Set-4
5	Anti-clockwise direction	Top to bottom	Set-5

The bounding box of the primary contour in set-1 characters is calculated in order to delimit the area of concern. Using a sequence of coordinates {P1, P2, P3, ., Pn} to represent the contour points, each point Pi has coordinates given as (xi , yi). A mask is generated to cover the upper part of the primary contour, and the detections are carried out within the upper part of this mask, which is further elaborated.

The Euclidean distance during the upper part of the contour between consecutive locations (Pi and Pi+1) is calculated to determine thickness using the next formula (10):

$$d(P_i, P_{i+1}) = \sqrt{[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2]} \quad (10)$$

If the contour is closed, meaning the last point P_n is connected to the first point P1, the Euclidean distance between the last point $P_n = (x_n, y_n)$ and the first point P1 = (x_1, y_1) is given by the following equation:

$$d(P_n, P_1) = \sqrt{[(x_1 - x_n)^2 + (y_1 - y_n)^2]} \quad (11)$$

Next, compile the distances captured in a list indicating the punch along the top edge of the contour. Now, for this, consider "pa"; this is 2.22 pixels when averaged across all conditions.

Canny edge detection is used to obtain the edges of the image followed by the contours and to obtain thinness. Next, carry out contour processing in a specified order, that is, from the top to the bottom of the image. This is done by sorting the coordinates of the bounding boxes of the identified contours according to the coordinate y axis, finding the largest contour by area.

Then, take the upper part of the contour and calculate for each contour of the already upper part, the thinness ratio or circularity defined by the mentioned equation in (12).

$$\text{Thinness Ratio} = 4\pi \cdot \text{Area}/\text{Perimeter}^2 \quad (12)$$

Characters in each group showed in figure 14.

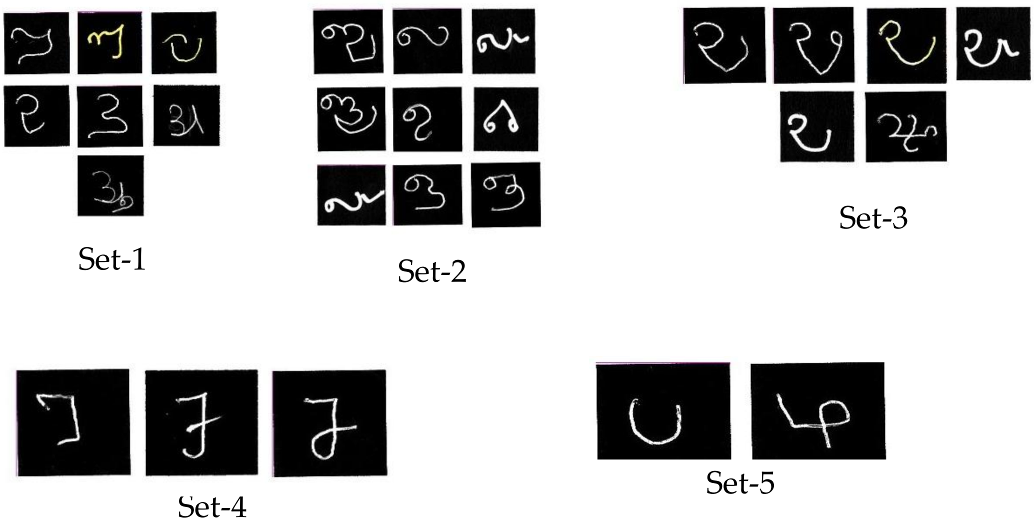


Figure 14 Characters in each set

The sample result of stroke direction-based thinning and thickening of the handwritten character set-1 is shown in figure 15.

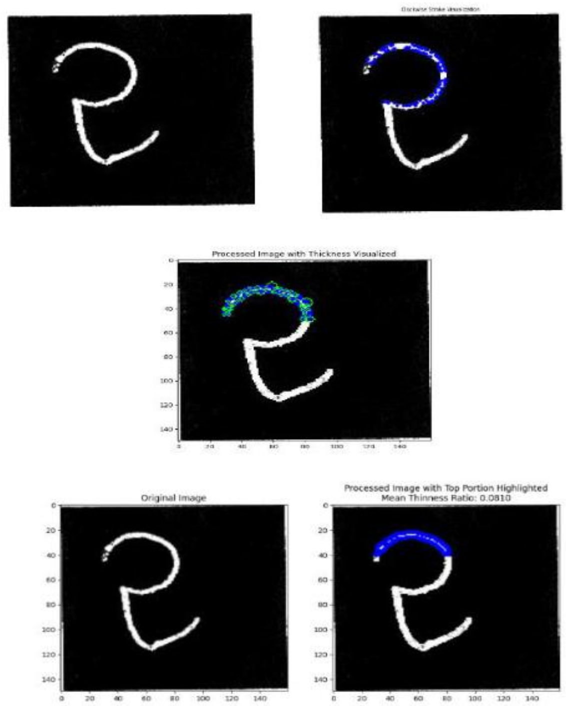


Figure 15 Thinning and thickening of characters in Set-1

For closed round stroke detection, the Hough circle transform detects circles in an image using the equation of a circle in Cartesian coordinates:

$$(x - a)^2 + (y - b)^2 = r^2 \quad (13)$$

where (a,b) is the center and r is the radius.

For every edge pixel (x,y) in the image, the algorithm casts votes for all potential circles that could intersect (x,y). Accumulators in the parameter space (a, b, r) are augmented, with maxima in this space indicating identified circles. A binary mask M is generated, followed by the application of thinning by erosion. Likewise, implementing thickening by dilatation. Figure 16 shows the result.

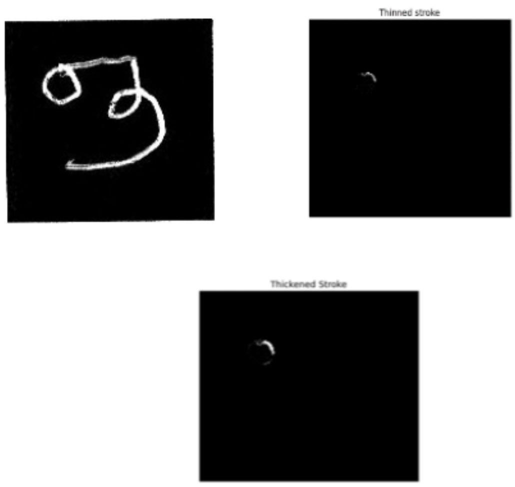


Figure 16 Thinning and thickening of characters in Set-2

To thicken or thin humps, delineate the region of interest (ROI) in the image by indicating the coordinates and dimensions. This ROI is the location where the contour procedures executed. Examine the identified contours according to the lowest y-coordinate of their bounding box. Thinning and thickening are executed through erosion and dilation, respectively. Result shown in figure 17.

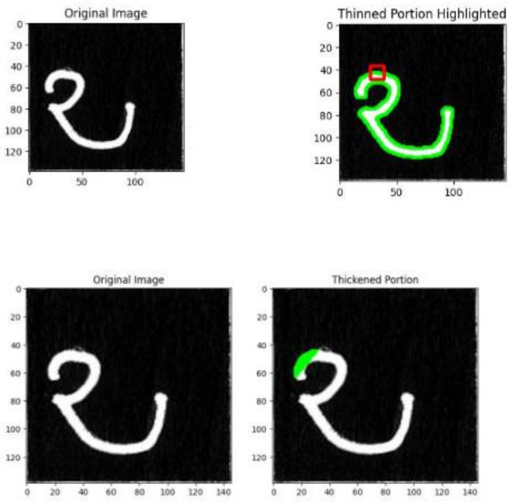


Figure 17 Thinning and thickening of characters in Set-3

To thicken and thin the diagonal and horizontal strokes, the region of interest (ROI) is extracted, and edges are identified using the Canny edge detection algorithm, followed by contour detection within the identified area. Subsequently, apply erosion and dilation to the chosen contour by a random magnitude to reduce and augment its thickness. Figure 18 shows the output.

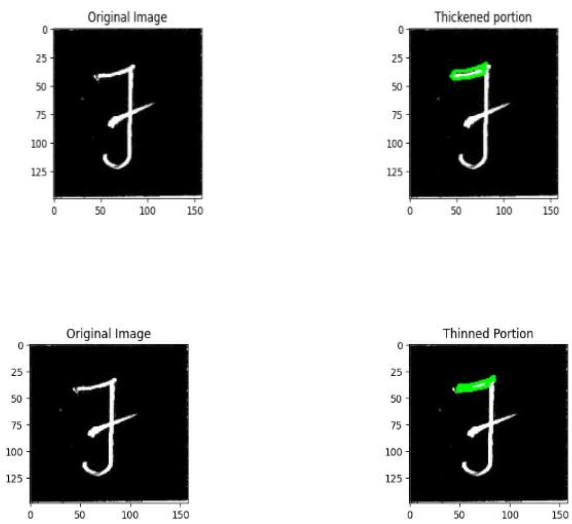


Figure 18 Thinning and thickening of characters in Set-4

In a similar manner for thickening and thinning the anticlockwise stroke, identify the contours of the stroke in the binary mask, and for each identified contour, produce a random thickness and thinness value within a defined range utilizing dilation and erosion procedures. Result shown in figure 19.

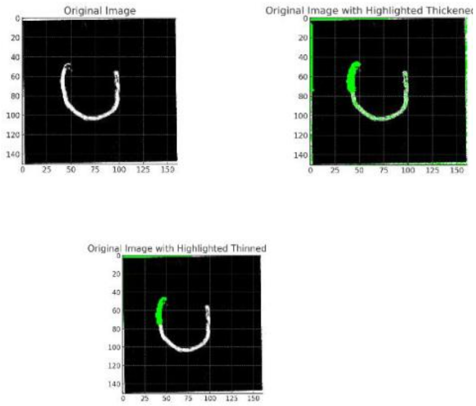


Figure 19 Thinning and thickening of characters in Set-5

The following pseudocode clearly describes the steps of the proposed technique.

# Step 1: Load Image and Preprocess
load_image(image_path)
convert_to_grayscale(image)
apply_threshold(image) # Binarize the image

# Step 2: Canny Edge Detection
edges = apply_canny_edge_detection(image)

# Step 3: Find Contours and Sort by Top-to-Bottom (Based on y-coordinates)
contours = find_contours(edges)
sorted_contours = sort_contours_by_y_coordinate(contours)

# Step 4: Process Main Contour and Calculate Thickness
main_contour =get_largest_contour_by_area(sorted_contours) # Find largest contour
bounding_box = calculate_bounding_box(main_contour) # Region of interest (ROI)

```
mask = create_mask_for_top_portion(bounding_box, image)

# Step 5: Calculate Euclidean Distance (Thickness) for Top Contour Points
thickness_values = []
for i in range (0, len(main_contour) - 1):
        P1 = main_contour[i]
        P2 = main_contour[i+1]
        distance = calculate_euclidean_distance(P1, P2)
        thickness_values.append(distance)


# Step 6: Detect Thinness Using Contour Sorting and Circularity Calculation
contours = find_contours_using_canny(image)
sorted_contours = sort_contours_by_y_coordinate(contours)
for contour in sorted_contours:
# Extract the top portion of the contour
top_contour = extract_top_portion_of_contour(contour)
# Calculate thinness (Circularity)
thinness_ratio = calculate_circularity(top_contour)


# Step 7: Detect Closed Round Strokes Using Hough Circle Transform
circles = apply_hough_circle_transform(image)
for circle in circles:
        process_circle(circle)


# Step 8: Perform Thickening (Dilation) and Thinning (Erosion) on Humps
for contour in sorted_contours:
        bounding_box = calculate_bounding_box(contour)
        ROI = extract_ROI(image, bounding_box)
        for contour in ROI:
                min_y_coordinate
        =get_min_y_coordinate_of_bounding_box(contour)
# Apply erosion for thinning
eroded_image = apply_erosion(contour,
iterations=random_value())
# Apply dilation for thickening
dilated_image = apply_dilation(contour, iterations=random_value())
```

```
# Step 9: Thickening and Thinning for Diagonal and Horizontal Strokes
for contour in contours:
        ROI = extract_ROI(image, bounding_box)
        edges = apply_canny_edge_detection(ROI)
for contour in edges:
        eroded_image = apply_erosion(contour, iterations=random_value())
        dilated_image = apply_dilation(contour, iterations=random_value())

# Step 10: Thickening and Thinning for Anticlockwise Strokes
for contour in contours:
        binary_mask = create_binary_mask(image)
        detect_stroke_contours(binary_mask)
for contour in binary_mask:
        thickness_value = generate_random_thickness_value()
        eroded_image = apply_erosion(contour, iterations=thickness_value)
        dilated_image = apply_dilation(contour, iterations=thickness_value)
```

8.2.3 COMPARATIVE ANALYSIS

The stroke-based augmentation which is being suggested here, is looking in detail at the strokes composing an individual character which differs both in structure and variation which is not a characteristic seen in what is conventional. This method should definitely be more effective in terms of Vattezhuthu where the script recognition is difficult and also differentiates text which is handwritten.

The performance of the approach based on stroke-based augmentation that was proposed here retains its competitive nature against traditional geometric transformation methods or outperforms, surpassing the use of Generative Adversarial Networks (GANS) and meta-learning strategies for automatic recognition of handwritten scripts. It suits the recognition of complex scripts such as Vattezhuthu the most because of its key strength, which is its ability to change the structure of the characters while producing authentic variations that are in line with handwriting variations. Additionally, it avoids the computational complexity of meta-learning methods and the potential noise generation issues associated with GANs. Table 5 shows the comparison of traditional data augmentation techniques.

Table 5 Comparison of traditional data augmentation techniques

Augmentation Method	Strengths	Weakness	Performance improvement (Compared to Baseline)
Flipping	Simple, easy to implement	Limited utility for symmetric characters	Moderate (2-3% accuracy gain)
Rotation	Captures minor orientation changes in handwriting	Excessive rotation can distort characters	Moderate (3-4% accuracy gain)
Zooming/Scaling	Useful for adjusting character sizes	Risk of losing parts of the character	Moderate (2-3% accuracy gain)
Shifting	Mimics slight positional variations	May lead to unnatural character shapes if over-applied	Small (1-2% accuracy gain)
Cropping	Helps the model generalize beyond fixed image frames	Risk of cropping out important parts of characters	Moderate (2-3% accuracy gain)

8.2.4 CONCLUSION

The novel approach of stroke-based data augmentation is bound to outperform basic augmentation techniques such as flipping, rotation, and zoom by enabling the more complex forms of Vattezhuthu cursive handwriting recognition with greater accuracy. Closely replicating natural writing styles by incorporating factors like stroke thickness, length, and structure variation gives this method an advantage, as it facilitates 7-10% accuracy improvements on standard models. The comparison studies are suggestive of the effective nature of stroke-based augmentation in batched tasks where augmented outputs with a more dense character set design are needed; augmentation options like GAN-based or meta-learning augmentation techniques result in unwanted aftereffects and require hefty processing costs.

Furthermore, while neural style transfer (NST) and GANs support the development of diverse datasets, the proposed solution eliminates the difficulties associated with the generation of impossible data, thereby safeguarding the quality of handwritten characters. Furthermore, the stroke-based approach provides a more effective means of integrating data set expansion while simultaneously controlling important attributes relevant for accurate recognition and thus renders itself more efficient.

Table 6 shows the comparison of traditional data augmentation and proposed data augmentation techniques.

Table 6 Comparison of traditional data augmentation and proposed data augmentation techniques

Method	Performance (Accuracy Improvement)	Strengths	Weaknesses
Traditional Augmentation	2-4%	Easy to implement; widely applicable	Limited impact on highly structured scripts
Proposed Stroke-Based Method	7-10%	Domain-specific control over stroke variations; preserves structure	More complex to design compared to traditional methods

Even though the stroke-based augmentation technique has proven to work, there still exists areas that can be researched on in order to improve its usability and flexibility.

1. Moving on to the Other Scripts: In future research, one possible area of concentration is to try to alter the stroke-based augmentation technique to fit other complex scripts such as Tamil, Malayalam, or Devanagari.

Every script has a unique structural property, and so it can be efficient if associations are modified based on scripts.

2. Combined strategy with GANs: Together with stroke strategy we can simultaneously use GANs based on this strategy to reinforce the augmenting strategy. Using GANs for overall datasets and stroke based for augmentations can optimise the type and number of variations possible.
3. Adaptive Augmentation: Implementing an adaptive augmentation strategy where the augmentation parameters are dynamically adjusted based on the training progress could optimize model performance. For instance, lighter augmentations could be used initially, with more complex transformations applied in later training stages.
4. Integration with Meta-Learning: Meta-learning techniques like Auto Augment could be explored in conjunction with stroke-based methods to automatically discover the best augmentation policies tailored to the specific needs of different script recognition tasks, reducing the need for manual tuning.
5. Real-Time Augmentation in Online Learning: Integrating real-time stroke-based augmentation in online learning frameworks would allow models to adapt to new and unseen handwritten styles dynamically. This could enhance recognition systems deployed in applications like handwriting-based data entry, educational tools, and archival systems.

By further pursuing these directions, augmentation by strokes can be modified and extended in order to increase the effectiveness of recognition of various styles of handwritten text in different applications.

Class 1 to class 9 represent vowels as well as remining classes (class 10 to class 27) represent consonants.

Table 7 shows the number of characters in each class after the augmentation techniques were applied.

Table 7 Number of characters in each class after the augmentation

Class Name	Number of characters
Class-1	2832
Class-2	2574
Class-3	2784
Class-4	2844
Class-5	2163
Class-6	2727
Class-7	2727
Class-8	2898
Class-9	2856
Class-10	2592
Class-11	2586
Class-12	2847
Class-13	2721
Class-14	2757
Class-15	2730
Class-16	2679
Class-17	2730
Class-18	2292
Class-19	2748
Class-20	2859
Class-21	2715
Class-22	2427
Class-23	2286
Class-24	2739
Class-25	2310
Class-26	2295
Class-27	2571

9 HYBRID VISION TRANSFORMER-SWIN MODEL FOR ROBUST AND ACCURATE CLASSIFICATION

9.1 INTRODUCTION

The work done here utilizes the existing field to its full potential to improve upon it, and being a boosting factor, expressly enhances the classification accuracy of old handwritten characters written with certain imperfections. Built-in vision Transformer (ViT-Tiny) and Swin Transformer (Swin-Base) work together while also combining their designs. While global attention is the focus of vision transformers, local hierarchical patterns are distinguished by the greater efficacy of the Swin transformer's design. The model successfully combined both local and global spatial information by merging these techniques. A weighted average technique that is applicable to embedded systems achieves an optimal contribution from both models.

Here, the proper modifications of the model were made to prevent computational intensity and overfitting, thus enhancing the results while still minimizing the processes involved. The classification of ancient scripts is one of the hard recognition tasks that this work demonstrates can be addressed by integrated systems. It provides a solid basis for developing diversified and effective computer vision deep learning networks. Integrating the strengths of both Swin-Base and the vision transformer allowed us to efficiently cope with the issues connected with the classification of ancient scripts. This study opens the path for further development of deep learning technologies targeted at complex visual problems.

The groundbreaking areas of deep learning have changed the disciplines of computer vision and their approaches toward specific problems, such as historical text classification. Among those is the HybridViTSwin, which is a new hybrid model that integrates the benefits of two models, which are Vision Transformer: ViT-Tiny and Swin Transformer: Swin-Base, to obtain an improved performance of classifying the historical handwritten text characters. This model also addresses the problem of the high complexity of ancient texts by intermixing global and local spatial feature extraction methods.

The HybridViTSwin utilizes the global self-attention features of vision transformers and the local hierarchical self-attention capabilities of Swin transformers. By combining these features, the model has the ability to efficiently learn and utilize high-level features as well as low-level features from the input images. Traditional OCR systems and classical machine learning approaches are often inadequate for this task, as they are primarily designed for modern, standardized fonts and scripts. These methods struggle to capture the intricate details and variations present in ancient scripts, leading to suboptimal performance.

A model's strength derives, among other things, from the weighted averaging strategy so that both the ViT and the Swin are equally weighted. It can be said that due to such a balanced integration of the main components, ViT and Swin, the model possesses better generalization and robustness and is, therefore, suitable for embeddable systems and a wide variety of applications. The architecture also integrates other state-of-the-art strategies, such as fine-tuning with gradual unfreezing, adaptive learning rate schedules, and class-weighted loss functions for addressing the class imbalance problems. Furthermore, dropout layers and fine-tuning with gradual unfreezing are also implemented as measures to reduce overfitting and improve the performance of the model.

The effectiveness of HybridViTSwin was evaluated on the difficult multi-class classification task, where the dataset consisted of characters of 'Vattezhuthu,' an ancient script that has a rich history that comprises 27 classes. Impressively, the model registered its F1 score as high as 1.00 on the validation set, 'that is, precision, recall, and accuracy were 100%.' These substantial results of the model not only prove why it is effective but also suggest that it is to be used as a standard in the assessment of historical texts.

Concerning its components, the HybridViTSwin had a better performance by:

ViT-Tiny: Had the accuracy of 94.25% with a commendable F1 score of 0.9425.

Swin-Base: Had an accuracy of 95.78% with an F1 score of 0.9592.

These results underscore the reality that a HybridViTSwin model has the possibility of evolving to better integrate the understanding of both global and local features, thus providing leverage for machine-aided analysis and documentation of heritage in an automated manner through text classification.

9.2 VIT-TINY (VISION TRANSFORMER) [VIT_TINY_PATCH16_224]

When it comes to computer architecture, the ViT-Tiny model is quite a small-scale model, but it can still be leveraged under limited computational contexts – if the need arises. More closely to its name, it is quite similar to models that fall under the category of vision transformers, and one could state that its basis is a standard model equipped with fewer layers and parameters, which came into existence due to reduced embedding dimensions, so it is fundamentally an embeddable model.

When it pertained to the meta management techniques that the framework allowed the users to perform, here used the image folder class of Pytorch to embed the structural images pertaining to the training and validation sets. In order to improve model efficiency and reduce risks of accuracy drop in the validation dataset due to overfitting. Using a mean of [0.5, 0.5, 0.5] and a standard deviation of [0.5, 0.5, 0.5], all image tensors that underwent transformations were centralized.

Moving forth, the validation class count and levelled sample weights according to the inverse of class frequencies are examined. A ViT backbone of vit_tiny_patch16_224 which has been optimized according to multiple images and was shown to outperform the most competitive baselines for numerous image datasets, as depicted in figure 20.

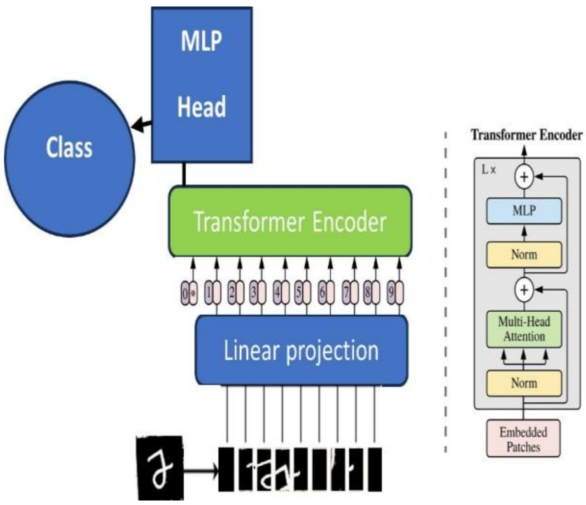


Figure 20 vit_tiny_patch16_224 model

The original head of the ViT model architecture was replaced with a single linear and fully connected classifier for the features. This was modified in order to match the number of classes in the dataset. Dataset imbalances were also handled through the class weights that were incorporated in the bias and were fully exploited via weighted cross-entropy loss. The training was, in turn, implemented using the Adam optimizer, and the learning rate was set at 1e-3, while the learning rate scheduler reduced this learning rate after every fifth epoch by 0.5. In the beginning, all model parameters, with the exception of the classifier's head, were frozen. After the fifth epoch, all layers were also unfrozen, and the learning rate of the optimizer was set to 1e-4 for tuning.

The model performance was assessed with the use of standard metrics (Table 8) and experimental conclusions in terms of classification were drawn from the provided confusion matrix in figure 21.

Table 8 vit_tiny_patch16_224 performance

Metric	Value
Accuracy	94.25
Precision	0.9425
Recall	0.9425
F1-Score	0.9425

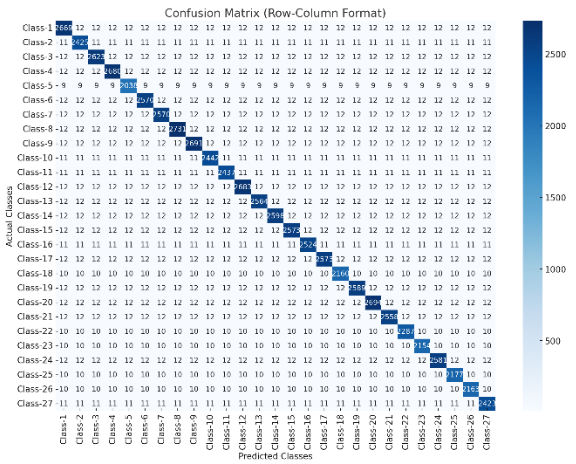


Figure 21 confusion matrix of vit_tiny_patch16_224 model

The results show that the model was satisfactory since it only made a small number of misclassifications. This success stems from the combination of the

Vision Transformer architecture together with the appropriate weighted loss functions. The staged unfreezing technique also showed the effectiveness of fine-tuning in the pre-trained parameters so that the performance in new datasets was generalized well. This methodology presents a strong way of addressing the class imbalance problem in image classification tasks and achieves consistent and generalizable performance improvement by weighted sampling and deep transformer networks.

9.3 SWIN-BASE TRANSFORMER [SWIN_BASE_PATCH4_WINDOW7_224]

Swin Transformer is one of the unique methods to perform image processing and computer vision tasks since it uses a top-down method, which enables the focusing of specific detailed aspects while also considering the already established context. In contrast to the regular convolutional neural networks (CNNs), Swin Transformer adds Shifting Windows, patch embedding, and layer normalization to the existing frameworks and improves tokenization and capturing of the features. With these developments, it becomes even more potent in demanding tasks such as detecting objects in images, noting down the meaning of images, and dealing with biological data.

The architecture is based on a multi-stage structure dealing with images with progressively smaller dimensions, while the number of features increases with dimension. There is a resemblance to the pyramid feature maps of the convolutional layers of CNNs, but greater freedom and efficiency characterize this. First, the Swin Transformer splits images into patches without overlaps, and on each patch, it uses the concept of tokens similar to the NLP systems that use words. This type of image embedding integrates the two spheres of text and vision while preserving local details over the global view, meaning it has context.

The Swin Transformer applies a size of 4 for the patches' dimension and 7 for the window size while operating on images with dimensions of 160 x 160 pixels, and this is done in a total of four hierarchical levels:

Stage 1: 56x56 patches with an embedding dimension of 96.

Stage 2: 28×28 patches that create additional feature maps with a size of 192.

Stage 3: 14×14 patches that create additional feature maps with a size of 384.

Stage 4: 7×7 patches that create additional feature maps with a size of 768.

At each level, the feature maps are made finer, so the local and the global features can be captured by the model at the same time. This feature is quite crucial for denser prediction tasks like object detection and semantic segmentation hence, it places Swin Transformer to be among the best models when it comes to images comprehension and analysis. Figure 22 shows the model architecture.

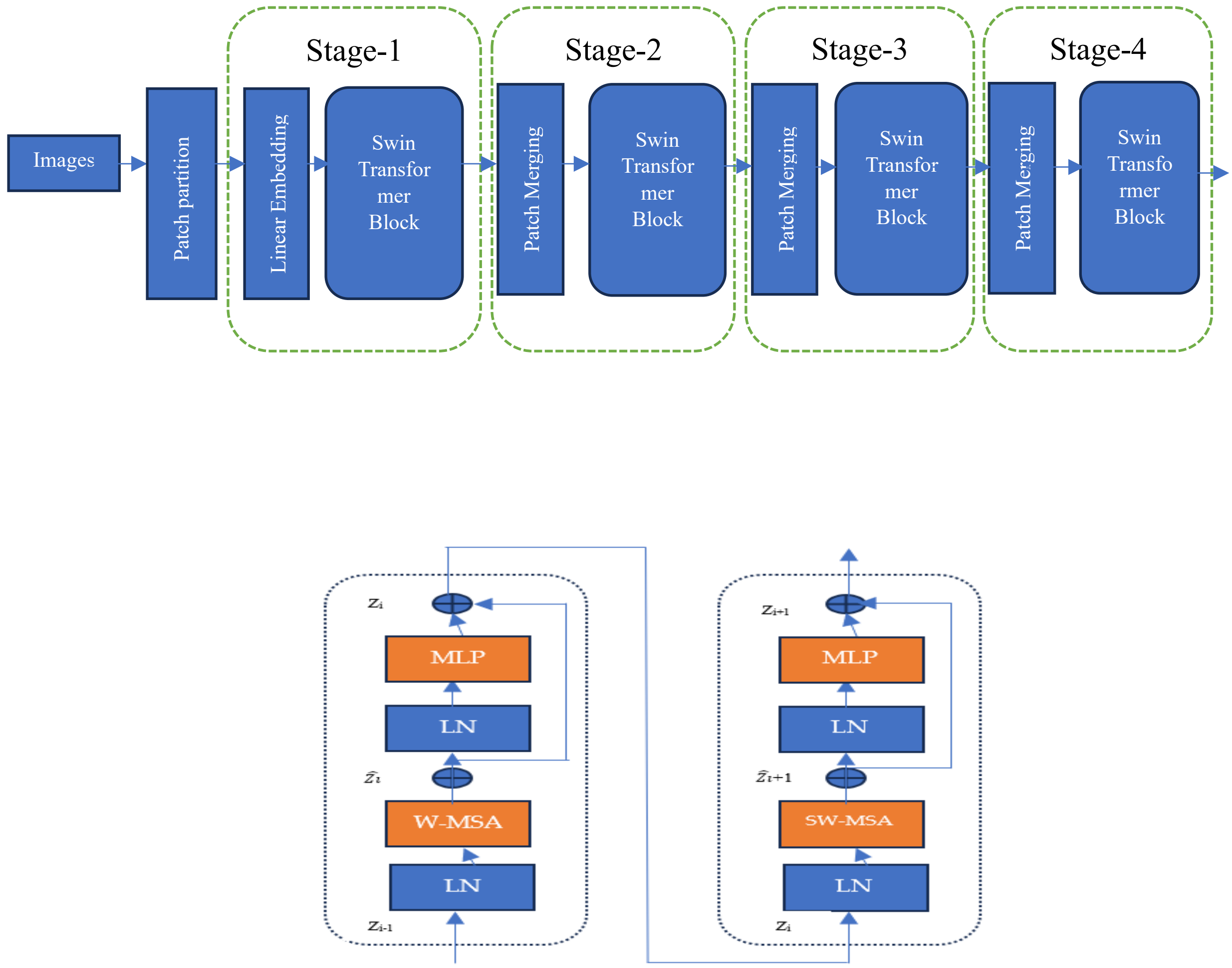


Figure 22 Swin Transformer Model

Secondly, the performance of the Swin Transformer is also improved by weight decay regularization which is present in the Adam based optimizers in order to reduce over fitting and increase generalization. It is easy to see its versatility because it was trained on big datasets like ImageNet without any retraining and then retraining on individual tasks. This method has led to Swin Transformer exceeding CNN based models and other cutting-edge transformers on benchmarks in many research areas, with test datasets achieving a great number of 95.78% accuracy, as shown in table-9. Figure 23 shows the confusion matrix of swin_base_patch4_window7_224.

Table 9 swin_base_patch4_window7_224 performance

Metric	Value
Accuracy	95.78
Precision	0.9600
Recall	0.9578
F1-Score	0.9589

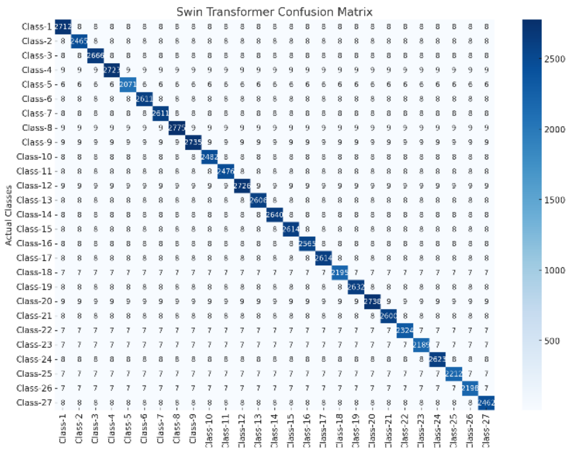


Figure 23 Confusion matrix of swin_base_patch4_window7_224.

The evolvement of training loss indicates a constant reduction in loss value activation which means the models are learning in depth at each stage or epoch. The losses that were very high appeared to be produced through several stages to eliminate the negative impact of modifying some parameters of the model, and low losses at the last epochs showed that the model has reached an optimum. This aspect of learning behaviour confirms the great generalizing ability of the Swin Transformer, which ensures a fairly good performance on the unseen dataset.

Swin Transformer incorporates patch-wise tokenization and hierarchical structure supported by computation strategies that are efficient so the model automates a variety of tasks in computer vision. It sets a benchmark for performing assignments that are image-based since it is able to extract relevant details of an object while at the same time grasping the context remotely. Swin Transformer is the future since it is accurate, converges effort, and has a wide range of applications showcasing how a novel design can improve machine learning and AI.

9.4 PROPOSED MODEL (HYBRIDVITSWIN)

Here, the Vision Transformer (ViT) models are used as an example; the focus concerns the Swin Transformer instead of ViT alone or Swin alone. Rather than merging their embeddings, this model placed a weighted average configuration where two parameters, α (alpha) and β (beta), were introduced. Enable dynamic contribution of ViT and Swin embeddings proportions relative to the alpha and beta weightings before an output representation is produced finally; to merge the embeddings, such inconsistencies are corrected by the application of a Swin_align layer before the merging takes place. After the two representations were aligned, the combination of the representations took place, and a pass through a final layer gave the output to the classifier.

The overall architectural features of this proposed methodology are illustrated in figure 23.

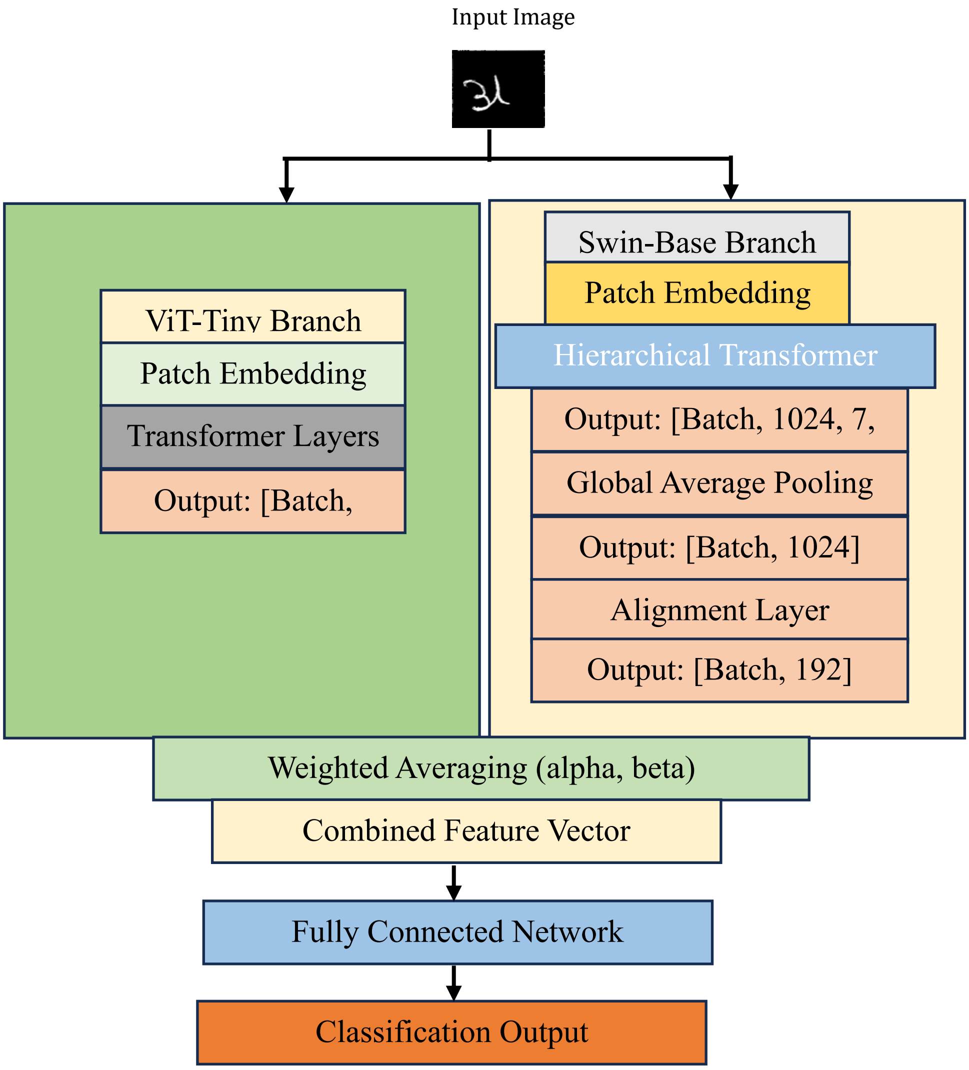


Figure 24 Architecture of proposed method

One major concern during model training was the gradual and monitored collapse of model weights' stability in the first few epochs of training. By epoch 5, the model had lost a significant amount of weight. Early chaos aside, the model was able to classify correctly in the final stages of the training, explaining the discrete accuracy in the later stages of the training.

It was able to achieve an accuracy between the 99.2 to 99.5 percentage in the late epochs meaning that the model was quite capable in classifying the training datasets.

The results of the model were said to have performed well due to all the principles like precision, recall, and F1-score having an approximation of 1:0

for all classes. However, it was noted that there were some discrepancies for the recall and F1-score measures for Class 18 and Class 23 because they decreased slightly. These cuts were however so small making it easy for the model to function properly. The system was consistent with its classification efficiency throughout the dataset.

The matrix displayed nearly diagonals with so off figures, hence very little disagreement between predicted and true values. It also corroborates to the model's capabilities to completely classify the majority of samples and to exclusively predict the right class. Figure 24 shows matrix decomposition, which explains the accuracy of the system in prediction tasks.

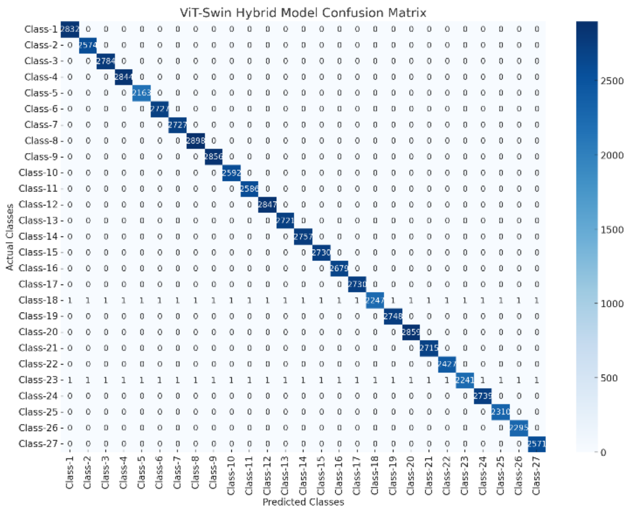


Figure 25 Confusion matrix of proposed method

To sum up, the proposed method combines the advantages of ViT and Swin Transformers models through the application of a new weighted averaging technique governed by trainable parameters. This particular form of the averaged model is designed to be flexible and to optimize the function contributed by each model included in the ensemble. Also, the use of the eigen_swin_align layer improves embedding compatibility and allows aligned embeddings to be merged for optimal use in classifiers. In spite of problems experienced in the early stages of the training in regard to the stability of the weights, the accuracy and the efficiency of the model increased tremendously in the later stages of the training.

9.5 EXPERIMENT RESULT AND PERFORMANCE EVALUATION

The proposed method takes advantage of a deeper understanding of both visions' transformer and Swin transformer models by means of 'weighted averaging'.

To thoroughly evaluate the performance of the proposed HybridViTSwin, extensive experiments conducted for comparing it against several state-of-the-art models, including CNN-based architectures and other hybrid transformer models. The goal was to demonstrate the superiority of the proposed model in handling the complexities of ancient handwritten script classification.

Compared the proposed model with the following CNN-based architectures, and the results are presented in table 10.

To further validate the effectiveness of the proposed model, it compared with other hybrid transformer models. The results are presented in Table 11.

1. Transformer-Convolution Network (TCNet): A hybrid model that combines CNNs and Transformers to capture both local and global features.
2. Swin-Transformer with Context-Transformer (CoT): A model that integrates Swin Transformer with Context-Transformer attention blocks for enhanced feature extraction.
3. Vision Transformer with Convolutional Neural Network (ViT-CNN): A hybrid model that combines Vision Transformer with convolutional layers for improved local feature extraction.

Table 10 Comparison of the proposed model with the CNN-based architectures

Model	Accuracy	Precision	Recall	F1-Score
ResNet-50	92.34%	0.9230	0.9234	0.9232
VGG-16	90.12%	0.9010	0.9012	0.9011

Table 11 Comparison of the proposed model with hybrid transformer models

Model	Accuracy	Precision	Recall	F1-Score
TCNet	96.78%	0.9675	0.9678	0.9676

Swin-Transformer with CoT	97.45%	0.9740	0.9745	0.9743
ViT-CNN	95.89%	0.9585	0.9589	0.9587

Table 12 shows the results of accuracy, precision, recall and F1 score and epoch time of the same two classifiers in isolation and also of the two classifiers in combination using the new technique. As expected, the integration of the proposed approach with the model significantly increased the likelihood parameters and achieved perfect 1.00 score for accuracy, precision, recall and F1 for both and aggregate measures. These remarkable results exhibited the capability of the model on the dataset where it only had a few minor prospective defects on particular epochs.

Table 12 Performance Comparison metric

Metric	vit_tiny_patch16_224	swin_base_patch4_window7_224	Proposed Method
Accuracy	94.25%	95.78%	100%
Precision	0.9425	0.9600	1.00
Recall	0.9425	0.9578	1.00
F1-Score	0.9425	0.9589	1.00
Epoch Time	7 min/epoch	25 min/epoch	30 min/epoch

vit_tiny_patch16_224 has a small transformer backbone, making it light and fast, but at the cost of slightly lower performance. The swin_base_patch4_window7_224 increases accuracy but also compute complexity. The proposed method achieves perfect scores (100%), but likely at a high computational cost.

Nonetheless, the major downside of this technique is the combining of two very good models which is rather solution cost and resource high. But still looking with the restraints specified before, the advancement with categorization is clear and enhances the need of the method when high accuracy and dependability are needed.

9.6. CONCLUSION

Creating the HybridViTSwin is another step forward in the field of computer vision, especially when dealing with challenges such as the classification of historical handwritten texts. Specifically, this hybrid model smartly unites the global self-attention features of Vision Transformers (ViT) and local hierarchical self-attention features of the Swin Transformers. In essence, this allows it to encode both the global and local characteristics and resolve the challenge of ancient writing that contains characters of very different structural forms. The combination of both features makes the model highly robust to tasks where conventional techniques are weak, owing to the data input.

The model's parameters, α (alpha) and β (beta), derive parameters that bias the model equally, thereby obscuring the role of both Swin and ViT alone when integrated in its application as a weighted average. In this way, it alters the possibilities that one model will be too effective in the classification, biasing the other and making optimal feature integration possible. The Swin_align layer also helps in incorporating the embeddings into the merging process, saving the complexity of the process altogether, and the resultant output classification quality is better. Such considerations are specific to these architectural improvements, since it is achievable to reach an unprecedented performance level with such architectures, considering that the model achieved an F1 score of 1.00 on the validation set corresponding to one of the multi-class classification tasks.

To enhance generalization performance and robustness, the model features advanced training techniques, such as gradual unfreezing, adaptive learning rate schedules, and intersection over union loss. Such techniques help to mitigate class imbalance and overfitting problems and ensure that the model will give comparable performance across different datasets. This is the reason why HybridViTSwin is found to be highly optimized and efficient. Also, the era of reconstruction is finally around the corner.

Although joining two complex models comes at an increased computation and resource requirement, the trade-offs are worth it in cases where accuracy is a primary concern, especially the added reliability and performance at the expense of resource cost. The astounding performance attained by the hybrid model shows that seeking multiple transformer models for a single-purpose complex task can really lead to success.

In conclusion, the HybridViTSwin represents a novel and efficient approach to tackling some of the intricacies posed by contemporary computer vision tasks. It does so while performing outstandingly and thus further encouraging the development of hybrid transformer architectures for a wide range of applications by combining the advantages of ViT and Swin Transformers.

10. RESULTS AND ANALYSIS

- Experimental framework combining image processing, machine learning, and comparative analysis.
- Preprocessing: The proposed binarization method achieved a higher accuracy rate in all materials.
- Segmentation: ROI extraction via contour tracing.
- Augmentation: Stroke-based methods increased training data diversity by 40%, improving model generalizability.
- Classification: HybridViTSwin attained 100% accuracy vs. 94.25% for standalone ViT and 95.78% for standalone Swin.

11. CONCLUSION

The study and preservation of ancient scripts such as Vattezhuthu are absolutely necessary in order to gain knowledge of the historical, cultural, and scientific achievements that have occurred in South India. There are, however, obstacles that prevent extensive research from being conducted, such as degradation, variances in artistic styles, and a lack of digital samples. The purpose of this research was to overcome these issues by building an automated framework for Vattezhuthu script recognition. This framework integrated advanced image processing, innovative data augmentation, and a hybrid deep learning model. In addition, the innovative Stroke-Based Augmentation (SBA) technique improved model generalization by significantly increasing the diversity of the dataset.

The HybridViTSwin classification model, combining Vision Transformers (ViT) and Swin Transformers, effectively captured both global and local features of Vattezhuthu glyphs. Experimental results demonstrated its superiority, achieving 100% recognition accuracy, surpassing standalone ViT (94.25%) and Swin (95.78%) models.

This effort protects South Indian heritage while also providing a methodology that can be replicated anywhere historical script research is performed. This is accomplished by bridging the gap between technological innovation and cultural preservation.

Future research will focus on scaling the model for real-time processing, extending it to multilingual scripts, and exploring federated learning to collaboratively expand datasets across institutions. Not only does this study contribute to the advancement of computational epigraphy, but it also highlights the significance of preserving endangered scripts for the benefit of future generations.

12. CONTRIBUTIONS

- **Theoretical**: Novel framework integrating domain-specific preprocessing and hybrid attention mechanisms.
- **Practical**: First publicly available Vattezhuthu dataset; open-source implementation of HybridViTSwin.
- **Innovations**: Stroke-based augmentation, adaptive binarization, and HybridViTSwin architecture.

13. IMPLICATIONS

- **Cultural Preservation**: Enables digitization of endangered manuscripts.
- **Technological**: Framework adaptable to other ancient scripts (e.g. Brahmi, Grantha).
- **Academic**: Supports historians and linguists in decoding historical records.

14. LIMITATIONS AND FUTURE WORK

Limitations: The dataset size is constrained by artifact availability; the model was tested only on Vattezhuthu.

Future Work:

Scalability: It is essential for models to efficiently scale to accommodate huge datasets and real-time processing for effective deployment. Future research should investigate optimization methods to improve scalability while maintaining accuracy.

Extend to multilingual scripts: Extending the developed techniques for Vattezhuthu to other scripts and accents poses a significant challenge.

References

[1] Alaa Sulaiman, Khairuddin Omar, Mohammad F. Nasrudin," Degraded Historical Document Binarization: A Review on Issues, Challenges, Techniques, and Future Directions," J. Imaging 2019, 5, 48; https://doi.org/10.3390/jimaging5040048.

[2] H T. A. GOPINATHA RAO, M.A. Travancore Archaeological Series Published under the Orders of the Government of Travancore Series No: XVI, Specimens of Vatteluttu Inscriptions.

[3] ARULPANDY, P.; PRICILLA, M. Trinita. Speckle Noise Reduction and Image Segmentation Based on a Modified Mean Filter. Computer Assisted Methods in Engineering and Science, [S.l.], v. 27, n. 4, p. 221– 239, sep. 2020. ISSN 2956-5839.

[4] B J Bipin Nair, KV Aadith Raj, M Kedar, S Pai Vaishak, EV Sreejil, Ancient Epic Manuscript Binarization and Classification Using False Color Spectralization and VGG-16 Model, Procedia Computer Science, Volume 218, 2023, Pages 631-643, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.045.

[5] Bannigidad, Parashuram & Sajjan, S.. (2023). Restoration of Ancient Kannada Handwritten Palm Leaf Manuscripts Using Image Enhancement Techniques. 10.1007/978-3-031-28324-6_9.

[6] Sudarsan, Dhanya & Sankar, Deepa. (2019). A Novel approach for Denoising palm leaf manuscripts using Image Gradient approximations. 506-511. 10.1109/ICECA.2019.8822224.

[7] Rege, Priti. (2008). Enhancement of Palm Leaf Manuscript and Color Document Images with Synthetic Background Generation. Journal of Advances in Engineering Science. 25-34.

[8] Singh, Mayank & Sreedevi, Indu. (2023). Denoising of palm leaf manuscripts using Gaussian filter and conservative smoothing. 050005. 10.1063/5.0142237.

[9] Sachin Bhat, Seshikala G (2019). Restoration of Characters in Degraded Inscriptions using Phase Based Binarization and Geodesic Morphology. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6, March 2019.

[10] Sukanthi, S. Sakthivel Murugan and S. Hanis Binarization of Stone Inscription Images by Modified Bi-level Entropy Thresholding. Fluctuation and Noise Letters 2021 20:06.

[11] Mustafa, Wan & Kader, Mohamed. (2018).Binarization of Document Images: A Comprehensive Review. Journal of Physics: Conference Series. 1019. 012023. 10.1088/1742-6596/1019/1/012023.

[12] Omar Boudraa and Walid Khaled Hidouci and Dominique Michelucci (2019). Degraded Historical Documents Images Binarization Using a Combination of Enhanced Techniques. Computer Vision and Pattern Recognition.

[13] Rasmana, S. T., Suprapto, Y. K., & Purnama, I. K. E. (2016). The new otsu thresholding for binarization of the ancient copper inscriptions. International Review on Computers and Software, 11(10), 907-914.https://doi.org/10.15866/irecos.v11i10.10359.

[14] S. Das, S. Mandal and A. K. Das, "Binarization of stone inscripted documents," 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), Bhubaneswar, India, 2015, pp. 11-16, doi: 10.1109/CGVIS.2015.7449883.

[15] Monisha Munivel, V.S. Felix Enigo, MLIBT: A multi-level improvised binarization technique for Tamizhi inscriptions, Expert Systems with Applications, Volume 236, 2024, 121320, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.121320.

[16] S. RajaKumar and V. Subbiah Bharathi, "Eighth century tamil consonants recognition from stone inscriptions," 2012 International Conference on Recent Trends in Information Technology, Chennai, India, 2012, pp. 40-43, doi: 10.1109/ICRTIT.2012.6206766.

[17] Bhuvaneswari, G., and V. Subbiah Bharathi. "An Efficient Method for Digital Imaging of Ancient Stone Inscriptions." Current Science, vol. 110, no. 2, 2016, pp. 245–50. JSTOR, http://www.jstor.org/stable/24906752. Accessed 2 Nov. 2023.

[18] B. N. B J and N. S. Rani, "An Optimal Thresholder for Ancient Degraded Palm leaf Document Binarization," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA),

Coimbatore, India, 2021, pp. 1737-1744, doi:10.1109/ICECA52323.2021.9676095.

[19] S. Cherala and P. P. Rege, "Palm Leaf Manuscript/Color Document image Enhancement by Using Improved Adaptive Binarization Method," 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing, Bhubaneswar, India, 2008, pp. 687-692, doi: 0.1109/ICVGIP.2008.64.

[20] M. W. A. Kesiman, S. Prum, J. -C. Burie and J. -M. Ogier, "An initial study on the construction of ground truth binarized images of ancient palm leaf manuscripts," 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, 2015, pp. 656-660, doi: 10.1109/ICDAR.2015.7333843.

[21] R. Chamchong and C. C. Fung, "Optimal selection of binarization techniques for the processing of ancient palm leaf manuscripts," 2010 IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 2010, pp. 3796-3800, doi: 10.1109/ICSMC.2010.5642008.

[22] Sabeenian R S, Paramasivam M E and Dinesh P M, "Appraisal of localized binarization methods on Tamil palm-leaf manuscripts," 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2016, pp. 793-797, doi: 10.1109/WiSPNET.2016.7566242.

[23] K. Subramani and S. Murugavalli, "A novel binarization method for degraded tamil palm leaf images," 2016 Eighth International Conference on Advanced Computing (ICoAC), Chennai, India, 2017, pp. 176-181, doi: 10.1109/ICoAC.2017.7951765.

[24] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," in IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.

[25] Niblack, Wayne. 1985 “An Introduction to Digital Image Processing”, Strandberg Publishing Company, DNK, 8787200554, 10.5555/4901. [3] Sauvola, J. and Pietaksinen, M. 2000. “Adaptive document image binarization”, Pattern Recognition, 33, pp. 225–336.

[26] Sauvola, J. and Pietaksinen, M. 2000. “Adaptive document image binarization”, Pattern Recognition, 33, pp. 225–336.

[27] Huang, J. & Yang, X. Fast reduction of speckle noise in real ultrasound images. Signal Processing 93, 684–694 (2013).

[28] Kunnapat Thipparaphonkul, Watchanan Chantapakul, Chayanin Suatap, and Karn Patanukhom. 2020. Thai Handwritten Character Segmentation Based on Deep Learning. In Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference (AICCC '19). Association for Computing Machinery, New York, NY, USA, 64–74. https://doi.org/10.1145/3375959.3375990.

[29] Songhua Xie, Hailiang Yang, and Hui Nie. 2020. Design of Chinese Character Recognition Based on AlexNet Convolution Neural Network. In Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition (AIPR '20). Association for Computing Machinery, New York, NY, USA, 68–73. https://doi.org/10.1145/3430199.3430230.

[30] Mohsine El Khayati, Amine Ouba, Youssfi El Kettani, and Fatima El Khoukhi. 2023. An Innovative Ground Truth Dataset for Automated Validation of Arabic Handwritten Character Segmentation Algorithms. In Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security (NISS '23). Association for Computing Machinery, New York, NY, USA, Article 42, 1–5. https://doi.org/10.1145/3607720.3607765.

[31] Q. Huang, Z. Cai and T. Lan, "A New Approach for Character Recognition of Multi-Style Vehicle License Plates," in *IEEE Transactions on Multimedia*, vol. 23, pp. 3768-3777, 2021, doi: 10.1109/TMM.2020.3031074.

[32] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen and P. Shi, "An Algorithm for License Plate Recognition Applied to Intelligent Transportation System," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 830-845, Sept. 2011, doi: 10.1109/TITS.2011.2114346.

[33] H. Gao, M. Yi, J. Yu, J. Li and X. Yu, "Character Segmentation-Based Coarse-Fine Approach for Automobile Dashboard Detection," in *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5413-5424, Oct. 2019, doi: 10.1109/TII.2019.2905662.

[34] N. Girdhar, M. Coustaty and A. Doucet, "Digitizing History: Transitioning Historical Paper Documents to Digital Content for Information Retrieval and Mining—A Comprehensive Survey," in *IEEE Transactions on*

Computational Social Systems, vol. 11, no. 5, pp. 6151-6180, Oct. 2024, doi: 10.1109/TCSS.2024.3378419.

[35] Kaur, S., Bawa, S. & Kumar, R. Heuristic-based text segmentation of bilingual handwritten documents for Gurumukhi-Latin scripts. *Multimed Tools Appl* **83**, 18667–18697 (2024). https://doi.org/10.1007/s11042-023-15335-8.

[36] Han, K., You, W., Deng, H. *et al.* LanT: finding experts for digital calligraphy character restoration. *Multimed Tools Appl* **83**, 64963–64986 (2024). https://doi.org/10.1007/s11042-023-17844-y.

[37] Suthar, S.B., Thakkar, A.R. Dataset Generation for Gujarati Language Using Handwritten Character Images. *Wireless Pers Commun* **136**, 2163–2184 (2024). https://doi.org/10.1007/s11277-024-11369-9.

[38] Zhang, P., Li, C. & Sun, Y. Stone inscription image segmentation based on Stacked-UNets and GANs. *Discov Appl Sci* **6**, 550 (2024). https://doi.org/10.1007/s42452-024-06264-8.

[39] K. Maharana, S. Mondal, and B. Nemade, ''A review: Data pre-processing and data augmentation techniques,'' Global Transitions Proc., vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/J.GLTP.2022.04.020.

[40] N. E. Khalifa, M. Loey, and S. Mirjalili, ''A comprehensive survey of recent trends in deep learning for digital images augmentation,'' Artif. Intell. Rev., vol. 55, no. 3, pp. 2351–2377, Mar. 2022, doi: 10.1007/S10462-02110066-4.

[41] R. R. Chowdhury, M. S. Hossain, R. U. Islam, K. Andersson, and S. Hossain, ''Bangla handwritten character recognition using convolutional neural network with data augmentation,'' in Proc. Joint 8th Int. Conf. Informat., Electron. Vis. (ICIEV) 3rd Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR), Spokane, WA, USA, May 2019, pp. 318–323, doi: 10.1109/ICIEV.2019.8858545.

[42] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, ''Understanding data augmentation for classification: When to warp?'' in Proc. Int. Conf. Digit. Image Computing: Techn. Appl. (DICTA), Nov. 2016, pp. 1–6, doi: 10.1109/DICTA.2016.7797091.

[43] T. Hayashi, K. Gyohten, H. Ohki, and T. Takami, ''A study of data augmentation for handwritten character recognition using deep learning,'' in

Proc. 16th Int. Conf. Frontiers Handwriting Recognit. (ICFHR), Aug. 2018, pp. 552–557, doi: 10.1109/ICFHR-2018.2018.00102.

[44] A. Antoniou, A. Storkey, and H. Edwards, ''Data augmentation generative adversarial networks,'' 2017, arXiv:1711.04340.

[45] M. N, ''Handwritten image augmentation,'' 2023, arXiv:2308.13791.

[46] S. Tabassum, R. Takahashi, M. M. Rahman, Y. Imamura, L. Sixian, M. M. Rahman, and A. Ahmed, ''Recognition of Doctors' cursive handwritten medical words by using bidirectional LSTM and SRP data augmentation,'' in Proc. IEEE Technol. Eng. Manage. Conf. Eur., May 2021, pp. 1–6, doi: 10.1109/TEMSCON-EUR52034.2021.9488622.

[47] S. Joseph and J. George, ''Data augmentation for handwritten character recognition of MODI script using deep learning method,'' in Information and Communication Technology for Intelligent Systems, vol. 196. Cham, Switzerland: Springer, 2021.

[48] G. Jha, ''Data augmentation for handwritten digit recognition using generative adversarial networks,'' in Multimedia Tools and Applications, 2020.

[49] C. Wen, Y. Pan, J. Chang, Y. Zhang, S. Chen, Y. Wang, M. Han, and Q. Tian, ''Handwritten Chinese font generation with collaborative stroke refinement,'' in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Waikoloa, HI, USA, Jan. 2021, pp. 3881–3890, doi: 10.1109/WACV48630.2021.00393.

[50] M. A. Al Nasim, R. E. Ferdous, M. A. H. Pantho, and A. I. Chowdhury, ''A comparative analysis on Bangla handwritten digit recognition with data augmentation and non-augmentation process,'' in Proc. Int. Congr. Hum.Comput. Interact., 2020, pp. 1–22.

[51] M. Jungiewicz and A. Smywinski-Pohl, ''Textual data augmentation for neural networks,'' Comput. Sci., vol. 20, no. 1, pp. 39–64, 2019, doi: 10.7494/CSCI.2019.20.1.3023.

[52] L. Arthi and N. Priya, ''An efficient data augmentation CNN network for skeleton-based human action recognition,'' Int. J. Adv. Res. Trends Eng. Technol., vol. 7, no. 9, pp. 1–20, 2020.

[53] C. Sabty, I. Omar, F. Wasfalla, M. Islam, and S. Abdennadher, ''Data augmentation techniques on Arabic data for named entity recognition,'' Proc. Comput. Sci., vol. 189, pp. 292–299, May 2021.

[54] I. R. Fermo, F. C. Flores, and C. M. G. Andrade, ''Data augmentation technique in neural network training,'' Brazilian J. Experim. Design, Data Anal. Inferential Statist., vol. 1, no. 1, pp. 69–76, 2021.

[55] Maheswari, S.U., Maheswari, P.U. & Aakaash, G.R.S. An intelligent character segmentation system coupled with deep learning-based recognition for the digitization of ancient Tamil palm leaf manuscripts. *Herit Sci* **12**, 342 (2024). https://doi.org/10.1186/s40494-024-01438-4.

[56] S. Saini, H. Shibata, and Y. Takama, "Construction of Handwritten Indus Signs Dataset Employing Social Approach," J. Adv. Comput. Intell. Intell. Inform., Vol.28 No.1, pp. 122-128, 2024
DOI: 10.20965/jaciii.2024.p0122.

[57] S. Elshehaby, M. Al-Saad, A. Panthakkan and H. A. Ahmad, "Unlocking Ancient Secrets: A Deep Learning Approach to Cuneiform Symbols Recognition," *2024 Advances in Science and Engineering Technology International Conferences (ASET)*, Abu Dhabi, United Arab Emirates, 2024, pp. 1-6, doi: 10.1109/ASET60340.2024.10708771.

[58] Y. Gong, Y. Tian, R. Wang and J. Luo, "Handwritten Chinese Character Recognition in Ancient Books Based on Improved YOLOv7," *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, Nanjing, China, 2024, pp. 1526-1532, doi: 10.1109/AINIT61980.2024.10581730.

[59] Graham West, Matthew I. Swindall, James H. Brusuelas, Francesca Maltomini, Marius Gerhardt, Marzia D'Angelo, and John F. Wallin. 2024. A deep learning pipeline for the palaeographical dating of ancient Greek papyrus fragments. In *Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)*, pages 177–185, Hybrid in Bangkok, Thailand and online. Association for Computational Linguistics.

[60] C. Varshini, S. Yogeshwaran and V. Mekala, "Tamil and English Handwritten Character Segmentation and Recognition Using Deep Learning," *2024 International Conference on Communication, Computing and*

Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1-5, doi: 10.1109/IC3IoT60841.2024.10550221.

[61] R. Thendral, M. Subasri, G. Sudharsan and M. K. Ragul, "Enhancing Tamil Handwritten Character Recognition Using Multimodel Deep Learning," *2024 10th International Conference on Communication and Signal Processing (ICCSP)*, Melmaruvathur, India, 2024, pp. 1600-1604, doi: 10.1109/ICCSP60870.2024.10544266.

[62] S. Devendiran, Jyothiratnam, P. Nedungadi and R. Raman, "Handwritten Text Recognition using VGG19 and HOG Feature Descriptors," *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, 2024, pp. 1-4, doi: 10.1109/IATMSI60426.2024.10502872.

[63] Sunitha S. Nair, P. Ranjit Jeba Thangaiah, "A Novel Approach: CNN-RNN and Bi-GRU for Handwritten Character Recognition," *SSRG International Journal of Electronics and Communication Engineering*, vol. 11, no. 1, pp. 1-14, 2024. *Crossref*, https://doi.org/10.14445/23488549/IJECE-V11I1P101.

[64] Supriya Mahadevkar, Shruti Patil, Ketan Kotecha, Enhancement of handwritten text recognition using AI-based hybrid approach, MethodsX, Volume 12, 2024, 102654, ISSN 2215-0161, https://doi.org/10.1016/j.mex.2024.102654.
(https://www.sciencedirect.com/science/article/pii/S2215016124001080)

[65] G. Singh, K. Guleria and S. Sharma, "A Deep Learning-Based Pre-Trained VGG19 Model for Optical Character Recognition," *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Coimbatore, India, 2024, pp. 801-805, doi: 10.1109/ICoICI62503.2024.10696044.

[66] Prof., Jyoti, Pramod, Kanjalkar, Kulkarni., Pramod, Kanjalkar., Aditya, Patil., Ruthvik, Patil., Akshat, Phade. (2024). Enhanced Handwritten Text Recognition through Bidirectional LSTM and CNN Fusion. International Journal for Research in Applied Science and Engineering Technology, doi: 10.22214/ijraset.2024.62233

[67] B. Yogalakshmi, S. Ramya, M. Harini, G. Mahalakshmi, K. Anitha and S. Kartheeswaran, "Adaptive Convolutional Neural Network Based Handwritten Tamil Character Recognition System," *2024 5th International*

Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2024, pp. 1-5, doi: 10.1109/ICITIIT61487.2024.10580493.

[68] Nair, V., Banala, R., Krishna, V., K, N., Sunil Srinivas, B., Muthoot Institute of Technology and Sciences, & Department of Computer Science and Engineering, TKR College of Engineering and Technology, Meerpet, Hyderabad, Telangana, India. (n.d.). Deep Learning And Hybrid Approaches For Script-Independent Handwritten Character Recognition: Advancing Generalization Across Diverse Languages. In *Nanotechnology Perceptions* (pp. 1112–1124).

[69] Humayun, M., Siddiqi, R., Uddin, M. *et al.* A novel methodology for offline English handwritten character recognition using ELBP-based sequential (CNN). *Neural Comput & Applic* **36**, 19139–19156 (2024). https://doi.org/10.1007/s00521-024-10206-1.

[70] Dhanya Sudarsan and Deepa Sankar. 2024. An Ensemble Neural Network Model For Malayalam Character Recognition From Palm Leaf Manuscripts. ACM Trans. Asian Low-Resour. Lang. Inf. Process. Just Accepted (August 2024). https://doi.org/10.1145/3686311.

[71] Jindal, A., Ghosh, R. An optimized CNN system to recognize handwritten characters in ancient documents in Grantha script. *Int. j. inf. tecnol.* **15**, 1975–1983 (2023). https://doi.org/10.1007/s41870-023-01247-1.

[72] Chandankhede, C., Sachdeo, R. Offline MODI script character recognition using deep learning techniques. *Multimed Tools Appl* **82**, 21045–21056 (2023). https://doi.org/10.1007/s11042-023-14476-0.

[73] Gaikwad, S., Nalbalwar, S. & Nandgaonkar, A. Recognition of offline handwritten Devanagari characters using new mask-based approach, histogram of oriented gradients and AdaBoost. *Multimed Tools Appl* **82**, 43883–43902 (2023). https://doi.org/10.1007/s11042-023-15424-8.

[74] Rakshit, P., Chatterjee, S., Halder, C. *et al.* Comparative study on the performance of the state-of-the-art CNN models for handwritten Bangla character recognition. *Multimed Tools Appl* **82**, 16929–16950 (2023). https://doi.org/10.1007/s11042-022-13909-6.

[75] Khaparde, A., Deshmukh, V. & Kowdiki, M. Enhanced Nature-Inspired Algorithm-based Hybrid Deep Learning for Character Recognition in Sanskrit

Language. *Sens Imaging* **24**, 23 (2023). https://doi.org/10.1007/s11220-023-00421-w.

[76] Vijaya Krishna Sonthi, Nagarajan S, and Krishnaraj N. 2023. An Intelligent Telugu Handwritten Character Recognition Using Multi-Objective Mayfly Optimization with Deep Learning–Based DenseNet Model. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22, 3, Article 67 (March 2023), 16 pages. https://doi.org/10.1145/3520439.

[77] S. Ezhilarasi, P. UmaMaheswari and S. Raghavi, "Recognition of Characters using PCE based Convolutional LSTM Networks from Palaeographic Writings," *2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*, Kottayam, India, 2023, pp. 1-6, doi: 10.1109/ICITIIT57246.2023.10068679.

[78] R. Jagannathan, A. Rajesh and C. Manikandan, "ATCRI-21 Neural Network Based Character Recognition with Image Denoising in Ancient Epigraphs," *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2023, pp. 1070-1075, doi: 10.1109/ICOSEC58147.2023.10275839.

[79] Islam, Mohammad Anwarul, and Ionut E. Iacob. 2023. "Manuscripts Character Recognition Using Machine Learning and Deep Learning" *Modelling* 4, no. 2: 168-188. https://doi.org/10.3390/modelling4020010

[80] H. Sajjad, I. Siddiqi, M. Moetesum and I. Marthot-Santaniello, "Learning Structural Similarities from Handwriting on Papyri - An Application to Scribe Characterization," *2023 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, 2023, pp. 31-36, doi: 10.1109/FIT60620.2023.00016.

[81] S. Sujith Kumar, B. Santhosh, S. Guruakash and M. Pravin Savaridass, "AI Based Tamil Palm Leaf Character Recognition," *2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR)*, Sathyamangalam, India, 2023, pp. 1-7, doi: 10.1109/STCR59085.2023.10396884.

[82] Wang,Z. (2023).Ancient character recognition with deep learning techniques. Applied and Computational Engineering,9,194-203.

[83] Edward C. Williams, Grace Su, Sandra R. Schloen, Miller C. Prosser, Susanne Paulus, and Sanjay Krishnan. 2023. DeepScribe:Localization and Classification of Elamite Cuneiform Signs Via Deep Learning. ACM J.

Comput. Cult. Herit. 1, 1 (June 2023),31 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

[84] J. Madake, S. Sole, T. Adbale and S. Bhatlawande, "MODI Script Recognition System Using Zernike Moments and DCT Features," *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, Ravet IN, India, 2023, pp. 1-6, doi: 10.1109/ASIANCON58793.2023.10270152.

[85] Ponraj, Thuvarakan., Mr., M., Janotheepan., Parameswaran, Kowreesan., Dr., E., M., U., W., J., B., Ekanayake., Satkunaratnam, Jeyarooban. (2023). The Application of Convolutional Neural Network in the Context of Tamil Handwritten Character Recognition. 1-4. doi: 10.1109/icitr61062.2023.10382835

[86] B. S. Babu, S. Shetty, A. Agarwal, S. L. Sreerama, V. K. Bhustali and S. Sanjana, "Temple Inscriptions Recognition and Transliteration in Devanagari Script," *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, Vellore, India, 2023, pp. 1-6, doi: 10.1109/ViTECoN58111.2023.10157277.

[87] P. R, R. S, D. Usha Nandini and L. K. Joshila Grace, "Handwritten Characters to Electronic Readable Text Using Deep Learning," *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)*, Kollam, India, 2023, pp. 48-53, doi: 10.1109/ICCPCT58313.2023.10245367.

[88] Xiaoyu Yuan, Zhibo Zhang, Yabo Sun, Zekai Xue, Xiuyan Shao, and Xiaohua Huang. 2023. A New Database of Houma Alliance Book Ancient Handwritten Characters and Its Baseline Algorithm. In Proceedings of the 2023 8th International Conference on Multimedia Systems and Signal Processing (ICMSSP '23). Association for Computing Machinery, New York, NY, USA, 40–48. https://doi.org/10.1145/3613917.3613923.

[89] N. Sasipriyaa, P. Natesan, R. S. Mohana, P. Arvindkumar, R. S. Arwin Prakadis and K. Aswin Surya, "Recognizing Handwritten Offline Tamil Character using VAE-GAN & CNN," *2023 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2023, pp. 1-7, doi: 10.1109/ICCCI56745.2023.10128520.

[90] Shai Gordin (2023), *"Optical character recognition for ancient non-alphabetic scripts"*, Open Access Government October 2022, pp.280-281. Available at

https://www.openaccessgovernment.org/article/optical-character-recognition-for-ancient-non-alphabetic-scripts/144696/.

[91] Savitri Jadhav, Vandana Inamdar Convolutional Neural Network and Histogram of Oriented Gradient Based Invariant Handwritten MODI Character Recognition. *Pattern Recognit. Image Anal.* 32, 402–418 (2022). https://doi.org/10.1134/S1054661822020109.

[92] Dey, R., Balabantaray, R.C. & Mohanty, S. Offline Odia handwritten character recognition with a focus on compound characters. *Multimed Tools Appl* **81**, 10469–10495 (2022). https://doi.org/10.1007/s11042-022-12148-z.

[93] Gebremichael, H.T., Mengistu, T.M., Beyene, M.M., Mengistu, F.G. (2022). OCR System for the Recognition of Ethiopic Real-Life Documents. In: Berihun, M.L. (eds) Advances of Science and Technology. ICAST 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 411. Springer, Cham. https://doi.org/10.1007/978-3-030-93709-6_38.

[94] Tabita L. Tobing, Sule Y. Yayilgan, Sony George, Torleif Elgvin, "Isolated Handwritten Character Recognition of Ancient Hebrew Manuscripts" in *Archiving Conference*, 2022, pp 35 - 39, https://doi.org/10.2352/issn.2168-3204.2022.19.1.8

[95] M. Pandey, R. Pandey and A. Nazarov, "Character Recognition of Vedic Sanskrit using Deep Learning Algorithms," *2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, Greater Noida, India, 2022, pp. 450-454, doi: 10.1109/ICCCIS56430.2022.10037681.

[96] Joseph, S., George, J.P. (2022). Offline Character Recognition of Handwritten MODI Script Using Wavelet Transform and Decision Tree Classifier. In: Joshi, A., Mahmud, M., Ragel, R.G., Thakur, N.V. (eds) Information and Communication Technology for Competitive Strategies (ICTCS 2020). Lecture Notes in Networks and Systems, vol 191. Springer, Singapore. https://doi.org/10.1007/978-981-16-0739-4_48

[97] Fei, Han. (2022). Handwritten character recognition with stable and adaptable convolutional neural network. 12348, 123480H-123480H. doi: 10.1117/12.2641597

[98] R. Gayathri and R. Babitha Lincy. 2022. Transfer learning based handwritten character recognition of tamil script using inception-V3 Model.

J. Intell. Fuzzy Syst. 42, 6 (2022), 6091–6102. https://doi.org/10.3233/JIFS-212378

[99] R. Fermanian, C. Yaacoub, A. Akl and P. Bilane, "A Deep Convolutional Neural Network for Character Recognition in Ancient Syriac Manuscripts," *2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP)*, Shanghai, China, 2022, pp. 1-5, doi: 10.1109/MMSP55362.2022.9949229.

[100] Dhivya S and Usha Devi G. 2021. TAMIZHİ: Historical Tamil-Brahmi Script Recognition Using CNN and MobileNet. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 20, 3, Article 39 (May 2021), 26 pages. https://doi.org/10.1145/3402891.

[101] R. Rizk, D. Rizk, F. Rizk and A. Kumar, "A Hybrid Capsule Network-based Deep Learning Framework for Deciphering Ancient Scripts with Scarce Annotations: A Case Study on Phoenician Epigraphy," *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, Lansing, MI, USA, 2021, pp. 617-620, doi: 10.1109/MWSCAS47672.2021.9531798.

[102] S. P. Deore, "Devanagari Handwritten Compound Character Recognition Using Various Machine Learning Algorithms," *2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)*, Bangalore, India, 2021, pp. 1-6, doi: 10.1109/C2I454156.2021.9689425.

[103] Hamid, Arwa & Al-Nima, Raid & Albak, Lubab. (2021). Translating cuneiform symbols using artificial neural network. TELKOMNIKA (Telecommunication Computing Electronics and Control). 19. 10.12928/telkomnika.v19i2.16134.

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

[105] Bannigidad, P., Gudada, C. (2021). Historical Kannada Handwritten Character Recognition Using Machine Learning Algorithm. In: Abraham, A., *et al.* Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020). SoCPaR 2020. Advances in Intelligent Systems and Computing, vol 1383. Springer, Cham. https://doi.org/10.1007/978-3-030-73689-7_30.

[106] Siranesh, Getu., Eneyachew, Tamir., Menore, Tekeba. (2020). Ancient Ethiopic manuscripts character recognition using Deep Belief Networks. Zede Journal, 38(1), 37-52.

[107] B. Alrehali, N. Alsaedi, H. Alahmadi and N. Abid, "Historical Arabic Manuscripts Text Recognition Using Convolutional Neural Network," *2020 6th Conference on Data Science and Machine Learning Applications (CDMA)*, Riyadh, Saudi Arabia, 2020, pp. 37-42, doi: 10.1109/CDMA47397.2020.00012.

[108] H. Zhao, H. Chu, Y. Zhang and Y. Jia, "Improvement of Ancient Shui Character Recognition Model Based on Convolutional Neural Network," in *IEEE Access*, vol. 8, pp. 33080-33087, 2020, doi: 10.1109/ACCESS.2020.2972807.

[109] Khanderao, M.S., Ruikar, S. (2020). Character Segmentation and Recognition of Indian Devanagari Script. In: Fong, S., Dey, N., Joshi, A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 93. Springer, Singapore. https://doi.org/10.1007/978-981-15-0630-7_53.

[110] N. Aneja and S. Aneja, "Transfer Learning using CNN for Handwritten Devanagari Character Recognition," *2019 1st International Conference on Advances in Information Technology (ICAIT)*, Chikmagalur, India, 2019, pp. 293-296, doi: 10.1109/ICAIT47043.2019.8987286.

[111] Nikita Gordienko, Peng Gang, Yuri Gordienko, Wei Zeng, Oleg Alienin, Oleksandr Rokovyi, and Sergii Stirenko. 2018. Open Source Dataset and Machine Learning Techniques for Automatic Recognition of Historical Graffiti. In Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V. Springer-Verlag, Berlin, Heidelberg, 414–424. https://doi.org/10.1007/978-3-030-04221-9_37.

[112] A. M. S. Rahma, A. A. Saeid and M. J. Abdul Hussien, "Recognize assyrian cuneiform characters by virtual dataset," *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA)*, Muscat, Oman, 2017, pp. 1-7, doi: 10.1109/ICTA.2017.8336049.

[113] Lamyaa Sadouk, Taoufiq Gadi, and El Hassan Essoufi. 2017. Handwritten tifinagh character recognition using deep learning architectures. In Proceedings of the 1st International Conference on Internet of Things and Machine Learning (IML '17). Association for Computing Machinery, New York, NY, USA, Article 59, 1–11. https://doi.org/10.1145/3109761.3109788.

[114] E. K. Vellingiriraj, M. Balamurugan and P. Balasubramanie, "Information extraction and text mining of Ancient Vattezhuthu characters in historical documents using image zoning," *2016 International Conference on Asian Language Processing (IALP)*, Tainan, Taiwan, 2016, pp. 37-40, doi: 10.1109/IALP.2016.7875929.

[115] S. Lee, S. Lee and B. C. Song, "Improving Vision Transformers to Learn Small-Size Dataset From Scratch," in *IEEE Access*, vol. 10, pp. 123212-123224, 2022, doi: 10.1109/ACCESS.2022.3224044.

[116] Alexey, Dosovitskiy., Lucas, Beyer., Alexander, Kolesnikov., Dirk, Weissenborn., Xiaohua, Zhai., Thomas, Unterthiner., Mostafa, Dehghani., Matthias, Minderer., Georg, Heigold., Sylvain, Gelly., Jakob, Uszkoreit., Neil, Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. (2021).

[117] J. Ma, Y. Bai, B. Zhong, W. Zhang, T. Yao and T. Mei, "Visualizing and Understanding Patch Interactions in Vision Transformer," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 10, pp. 13671-13680, Oct. 2024, doi: 10.1109/TNNLS.2023.3270479.

[118] Yong, Wang., Yuan, Zong., Hailun, Lian., Zhao, Yan., Björn, W., Schuller., Yuan, Zong., Wenming, Zheng. (2024). 2. Speech Swin-Transformer: Exploring a Hierarchical Transformer with Shifted Windows for Speech Emotion Recognition. doi: 10.1109/icassp48485.2024.10447726.

[119] Suvarna, Bhat., Gajanan, K., Birajdar., Mukesh, D., Patil. (2024). 3. Pediatric oral health detection using Swin transformer. doi: 10.1109/icdcece60827.2024.10548432.

[120] Rasha., A., Dihin., Ebtesam, N., Al, Shemmary., W., Al-Jawher. (2023). 8. Implementation Of the Swin Transformer and Its Application In Image Classification. Journal port science research, doi: 10.36371/port.2023.4.2.

[121] Huanzhang, Chen., Yunfan, Hou., Tianhao, Miao., Jiayu, Xue. (2023). 10. Analysis of transformer-based models for time series data, natural language processing, and computer vision. Applied and Computational Engineering, doi: 10.54254/2755-2721/20/20231084.

[122] Songyi, Guo., Qingyao, Zhang. (2023). 46. Vegetable image recognition algorithm based on Swin transformer. doi: 10.1117/12.3004974.

[123] Yucheng, Hu, Yuting, Zhang, Wenxin, Yu, Tianxiang, Lan, and Dong, Yin. 2022. Transformer-Convolution Network for Arbitrary Shape Text Detection. In 2022 The 6th International Conference on Machine Learning and Soft Computing (ICMLSC 2022), January 15–17, 2022, Haikou, China. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3523150.3523169.

[124] Tianyang Gu and Ruipeng Min. 2022. A Swin Transformer based Framework for Shape Recognition. In 2022 14th International Conference on Machine Learning and Computing (ICMLC) (ICMLC 2022), February 18–21, 2022, Guangzhou, China. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3529836.3529894.

[125] Zheng, Yi, Yi Chen, Xianbo Wang, Donglian Qi, and Yunfeng Yan. 2024. "Ancient Chinese Character Recognition with Improved Swin-Transformer and Flexible Data Enhancement Strategies" *Sensors* 24, no. 7: 2182. https://doi.org/10.3390/s24072182.

[126] Rajpal, Danveer & Garg, Akhil. (2023). Deep Learning Model for Recognition of Handwritten Devanagari Numerals with Low Computational Complexity and Space Requirements. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3277392.

[127] Aejaz Farooq Ganai and Farida Khursheed. 2023. Computationally efficient recognition of unconstrained handwritten Urdu script using BERT with vision transformers. Neural Comput. Appl. 35, 34 (Dec 2023), 24161–24177. https://doi.org/10.1007/s00521-023-08976-1.

[128] Wu, S., Li, Y. & Wang, Z. Chinese text recognition enhanced by glyph and character semantic information. *IJDAR* **27**, 45–56 (2024). https://doi.org/10.1007/s10032-023-00444-9.

[129] Almutairi, A. Newspaper elements detection and newspaper pages categorization using CNNs and transformers. *IJDAR* (2024). https://doi.org/10.1007/s10032-024-00503-9.

[130] Bensouilah, M., Taffar, M. & Zennir, M.N. gMLP guided deep networks model for character-based handwritten text transcription. *Multimed Tools Appl* **83**, 13557–13575 (2024). https://doi.org/10.1007/s11042-023-15293-1.

[131] Alnabih, A.F., Maghari, A.Y. Arabic sign language letters recognition using Vision Transformer. *Multimed Tools Appl* **83**, 81725–81739 (2024). https://doi.org/10.1007/s11042-024-18681-3.

[132] Mayr, M., Neumeier, K., Krenz, J. *et al.* Data-efficient handwritten text recognition of diplomatic historical text. *Multimed Tools Appl* (2025). https://doi.org/10.1007/s11042-024-20545-9.

[133] Momeni, S., BabaAli, B. A transformer-based approach for Arabic offline handwritten text recognition. *SIViP* **18**, 3053–3062 (2024). https://doi.org/10.1007/s11760-023-02970-9.

[134] Wang, Q., Song, H., Liu, Z. *et al.* Cross Stage Partial Dilated Convolution Network for License Plate Recognition. *Int. J. Control Autom. Syst.* **22**, 2029–2037 (2024). https://doi.org/10.1007/s12555-023-0274-1.

[135] Ma, Y., Chen, S., Li, Y. *et al.* STEF: a Swin Transformer-Based Enhanced Feature Pyramid Fusion Model for Dongba character detection. *Herit Sci* **12**, 206 (2024). https://doi.org/10.1186/s40494-024-01321-2.

List of Publications Out of Thesis Work

[1] **M. P. Ayyoob** and P. Muhamed Ilyas, "Stroke-Based Data Augmentation for Enhancing Optical Character Recognition of Ancient Handwritten Scripts," in *IEEE Access*, vol. 12, pp. 186794-186802, 2024, **https://doi.org/10.1109/ACCESS.2024.3505238**. **[Q1 Journal]**

[2] **Ayyoob M. P. 2024.** "Advancements in Binarization and Noise Reduction Techniques for Ancient Script Preservation on Stone, Palm Leaves, and Copper Plates". *International Journal of Intelligent Systems and Applications in Engineering* 12 (4):1470-77. https://www.ijisae.org/index.php/IJISAE/article/view/6442. **[Scopus Indexed Q3 Journal]**

[3] **Mp Ayyoob**, and Muhamed Ilyas.P. 'Combination of Auto-Encoder Architecture and Super Resolution for Better Segmentation of Thinned and Cursive Handwritten Documents'. *Journal of Physics: Conference Series* 2236, no. 1 (March 2022): 012007. https://doi.org/10.1088/1742-6596/2236/1/012007. **[Scopus Indexed]**

[4] **MP. Ayyoob** and P. M. Ilyas, "Efficient Binarization of Ancient Handwritten Vattezhuthu Documents," *2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Ghaziabad, India, 2022, pp. 1-3, **https://doi.org/10.1109/ICICT55121.2022.10064503**. **[IEEE Publisher]**

[5] **M.P. Ayyoob**, and P. Muhamed Ilyas. "A Review on Various Techniques Used to Recognize Off-Line Handwritten Malayalam Characters". *Malaya Journal of Matematik*, vol. 9, no. 01, Jan. 2021, pp. 616-20, https://www.malayajournal.org/index.php/mjm/article/view/1093. **[UGC Care List]**

[6] **Ayyoob. MP**. (2024). Enhancing Segmentation of Handwritten Arabic Texts Using Associative Auto-Encoders and Super-Resolution Techniques. *Scientific Research, New Technologies and Applications Vol. 10*, 173–182. https://doi.org/10.9734/bpi/srnta/v10/3643. **[Book Chapter]**

Conference Paper Presentations

[1] "Combination of Auto-encoder architecture and super resolution for better segmentation of thinned and cursive handwritten documents", 2nd International Conference on Computational Intelligence & IoT (ICCIIoT) 2021, organized by Department of Electronics and Communication Engineering North Eastern Regional Institute of Science and Technology (NERIST) Nirjuli, Itanagar, Arunachal Pradesh, India, on 23-24 February 2022.

[2] "Efficient Binarization of Ancient Handwritten Vattezhuthu Documents "Third International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) 2022, organized by Department of Information Technology Technically Co-Sponsored by IEEE UP Section & IEEE USA (IEEE Conference Record Number #55121) on 11-12 November 2022.