JSRT JOURNAL

PAGES: 64-73
5/26/24

JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY(JSRT) VOLUME-2 ISSUE-5 MAY

Digitization Of Tulu Handwritten Scripts- A Literature Survey

Sandhya Bangera[1], Subrahmanya Bhat[2]

[1]Research Scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore, India. ORCID ID:0000-0003-2009-0288 E-mail: sandhyabangera.ccis@srinivasuniversity.edu.in

[2]Professor and Dean, Institute of Computer Science and Information Science, Srinivas University, Mangalore, India. ORCID ID: 0000-0003-2925-1834; E-mail: itsbhat@gmail.com

ABSTRACT

Applications for handwritten script digitization include automated form processing, archival preservation, document digitization, handwriting recognition software, and historical document research. The performance of the recognition algorithms, the complex nature of the script, and the handwriting quality all affect how accurate the digitization is. Significant research on handwritten text recognition in scripts like Latin, Chinese, Arabic, and Japanese is reported. Still in its infancy, the work described on handwritten Indic scripts necessitates extensive research in this area. This thorough analysis offers an open view of different feature extraction and classification methods for the offline and online identification of handwritten Indic scripts.

The primary objective of this paper is to explore the accuracy and efficiency of recognition systems by analyzing several state-of-the-art studies and comparing and evaluating various techniques for Tulu script handwritten character recognition. This paper aims to describe various advancements reported over the last few decades in the field of handwritten scripts recognition.

Keywords: Handwritten Script Recognition, Digitization, Deep Learning, CNN, RNN, SVM, KNN.

I.INTRODUCTION:

The process of converting handwritten text or characters into a digital version that can be processed, saved, and examined by computers is known as "handwritten script digitization." To properly identify and transcribe the handwritten material, this approach usually makes use of image processing techniques and machine learning algorithms. Digitizing handwritten scripts has been increasingly popular as a way to digitize handwritten language into digital format in recent years. This technology is widely used in many different domains, such as text recognition systems, historical document preservation, and document digitization. It provides an answer to the difficulties in managing and examining handwritten materials, including old scripts, old manuscripts, and handwritten notes (Xiao, G, et.al. (2022).[1]). Textual information may be extracted and interpreted from a range of handwritten sources thanks to handwritten script digitization, which makes use of cutting-edge algorithms and machine learning approaches. This so makes it possible to handle data faster and more efficiently, increases the accessibility of handwritten documents, and safeguards valuable historical and cultural treasures for next generations. Over time, there has been a substantial evolution in the research and development of instruments and procedures for digitizing handwritten scripts. Authorship verification and forensic analysis depend heavily on the digitization of handwritten scripts. It gives academics and researchers the ability to analyze a person's handwriting traits in order to identify the author and validate historical records. Ultimately, digitizing handwritten scripts has transformed the way we manage and use handwritten material for analysis, preservation, and study, and has become an essential tool in many professions. (Schaeffer, et.al. (2020).[2]). Recent advances in artificial intelligence and neural network technology have produced considerable improvements in the accuracy and efficiency of handwritten script digitization. These advancements have made it possible to identify a large variety of handwriting styles and have prompted the development of stronger digitizing technologies. The discipline of forensics has benefited significantly from the use of handwritten script digitization. It has played a pivotal role in solving criminal cases by giving forensic experts powerful tools for interpreting handwritten evidence. Investigators can help address legal issues by gathering important information from handwritten materials through the precise identification and analysis of handwriting features.

The Eighth Schedule of the Indian Constitution now acknowledges 22 languages as scheduled languages in addition to Hindi and English. Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, Tamil, Telugu, Urdu, Bodo, Santhali, Maithili, and Dogri are among the languages that are officially acknowledged at the national level (Sharma, R., et.al. (2020).[3]).

1.1 Offline Handwritten Script Recognition:

After handwriting is created, it is usually captured as static images through scanning or photography for offline handwritten script recognition. The captured images only contain the visual representation of the handwritten script; they do not contain any additional information regarding stroke order or trajectory. Offline recognition relies just on the visual look of the characters and might not have access to extra contextual information accessible in online handwriting. Recognition algorithms analyse these static images to recognise and interpret the handwritten script. Applications like postal address recognition, historical document analysis, and document digitization frequently employ it.

1.2 Online Handwritten Script Recognition:

Handwriting is recorded as digital data while it is being formed for online handwritten script recognition. Usually, this entails using touchscreens or stylus pens digital writing instruments. The digital data is processed by recognition algorithms to identify and interpret the handwritten writing. The data comprises information about the trajectory of the pen or stylus, the stroke sequence, and maybe extra information like pressure and tilt. Additional contextual data, such as timing and stroke order, can be obtained through online recognition, which could help to increase recognition accuracy. Applications like digital handwriting input for tablets and signature verification systems frequently use it.

1.3 Steps involved in handwritten script digitization:

1. Image Acquisition: To create digital images, handwritten documents or scripts are scanned or photographed with digital cameras.

2. Pre-processing: The acquired images could have distortions, noise, or other artifacts that compromise the digitization accuracy. To improve the quality of the images, pre-processing methods like deskewing (straightening skewed images), binarization (converting images to black and white), and normalization are used.
3. Feature Extraction: This stage involves taking the handwritten characters' shapes, edges, and contours and extracting them from the previously processed images. To represent the handwritten characters in a format that recognition algorithms can understand, feature extraction is essential.
4. Recognition: The handwritten characters are trained to be recognized and transcribed by machine learning algorithms, especially those that are based on deep learning techniques such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). With the use of these algorithms, the extracted features are mapped to the appropriate characters in a digital alphabet.
5. Post-processing: To improve accuracy and refine results, post-processing techniques can be used after recognition. Spell checking, context-based correction, and language model integration are a few examples of this.
6. Output: After the text has been digitized, it is exported in an appropriate digital format, like plain text, PDF, or an organized document format.

II. OBJECTIVES:

(1) To digitize handwritten tulu scripts.
(2) To study various methodologies used for digitizing in other languages.
(3) To compare the accuracy achieved by different algorithms/ models in digitizing tulu script.
(4) To find a research gap and propose a model to improve accuracy.
(5) SWOT analysis of digitizing the tulu handwritten scripts.

III. METHODOLOGY:

The review in this paper is done by revising and analysing many different sources like journals, papers, conference papers, books published, websites.

IV. REVIEW OF LITERATURE/ RELATED WORKS:

Researchers and academics worldwide now have easier access to rare and fragile manuscripts thanks to the digitization of handwritten scripts, which has enabled the establishment of digital archives of historical records. This has created new avenues for interdisciplinary study, and archaeology has advanced our knowledge of historical occurrences and cultural advancements.
Image classification and optical character recognition are two important areas of computer vision research. More researchers are drawn to the fields of deep learning and machine learning as these methodologies advance and promise to produce nearly human-perfect models. Character recognition models for modern languages are widely available. However, it is still challenging to read handwritten text in Indian scripts. It is even more complex for scripts with a large alpha syllabary and complex nature. Tigalari is widely used for Sanskrit, Tulu, and Malayalam language documentation in coastal Karnataka and Kerala. Building a database, creating a deep convolution neural network (DCNN)-based architecture, training the model with the data, and using a test set to identify text are the steps in the process. The suggested approach, which functions as the genesis model for the

script, achieves an astounding 98.55% accuracy rate in classifying 46 essential Tigalari characters (Bhat, S et.al. (2020).[4]).
A wealth of knowledge about prehistoric history, customs, and culture can be found in the writings on palm leaves. Since these scriptures are a vital part of our rich cultural legacy, we must do everything in our power to preserve them. Unfortunately, there are several challenges in extracting information from palm leaves, such as the large character set, noise, and the challenge of reading the old Malayalam script, which makes the process time-consuming. Pattern recognition in Malayalam handwriting is a challenging and emerging field. Our proposed system uses contrast-based adaptive binarization and convolutional neural networks to translate the old Malayalam scripts to their modern version, thereby avoiding feature extraction and streamlining the entire process of information extraction from old palm leaves (Sudarsan, D., et.al. (2018).[5]).

Numerous handwritten document applications, such as bank checks, mail addresses, legal documents, prescription drugs, etc., have led to the development of numerous HCR techniques that have demonstrated satisfactory performance. Nonetheless, character recognition from historical documents is still an active field of study. For Ancient Scripts, HCR becomes more difficult because of different types of noise, low intensity, discoloration, etc. Historical documents have recorded and preserved a vast amount of knowledge in a variety of fields, including literature, architecture, history and civilization, economic factors, medicine, etc. Promising results have not been obtained for HCR for Ancient Scripts, despite the development of several techniques (Babu, N., et.al. (2019).[6]).

The first successes with optical character recognition (OCR) for historical scripts have created a huge potential for "smart" historical script studies. A smart ancient script database is necessary for "smart" ancient script research. Intelligent ancient script database software must be able to identify every facet and degree large amounts of data every ancient script material in order to gather the required for this astute research. Thus, the other main requirement going forward is to innovate a new digitized ancient script data system, one that includes full-scale supplementation to include all available materials, as well as newly inputted image data, in addition to integrating OCR functionality into this software (Bhat, S,, et.al. (2022).[7]).
The task of standardizing optical character recognition (OCR) for handwritten text is being undertaken by numerous researchers. As a result of their small sample size, some scripts are not used for OCR, but in the past few decades, researchers and research institutions have made significant advancements in other scripts. Relatively few studies have been published that recognize continuous character texts of different levels of complexity; the majority of reports focus on isolated characters. Opportunities to enhance OCR in particular dimensions for various scripts are created by this (Kataria, B., et.al. (2018).[8]).
Massive amount of ancient documents are created by our ancestors. This all documents are published already on the digital libraries on single click in worldwide. Also this documents published in Paper Media. Transcription of this documents are either prepared by manually or by the help of digital technology. These all ancient Documents are already available in raw images which are very useful to investigate the ancient philosophies of any field. However these raw images to be transliterate in electronic format by adopting the some image processing algorithms for appropriate transcription of the ancient documents. The character recognition difficulty is meager high due to some causes. This Research Paper investigates the problems faced in digital transcription of ancient handwritten text documents. The Research paper is organized in Three Section the First Section highlighted the Introduction about ancient script like MODI script, Handwritten Character Recognition techniques. The Second Section is all about the Problems investigating of MODI Character and last section is highlighted the Concluding remarks of analytical research (Tajne, N., , et.al. (2020).[9]).

Table 1: review of handwritten script recognition

Sl. No	Field and Focus of Research	Outcome	Reference
1.	a review of the different neural network methods available for multilingual script identification and recognition aids in gathering additional data for creating generic OCRs	helps gather more data for creating generic optical character recognition	Sinwar, D. et al. (2021)[10]
2.	The handwritten Devanagari script's features are extracted using a neural network diagonally based feature extraction method.	testing on more than a thousand characters, the offline Devanagari system's accuracy was found to be 85.78% match and 13.35% disagreement.	Agnihotri, V. P. et al. (2012)[11]
3.	Research areas in Indian offline handwritten character, digit, and word recognition are covered in the	broad survey report on developments in offline handwriting recognition for Indic script, with an emphasis on the last 15 years'	Singh, S. et al. (2023) [12]

	survey.	state-of-the-art work,	
4.	Off-line handwritten Devnagari character recognition using directional information derived from the gradient's arc tangent.	employed a 5-fold cross-validation scheme to test the system using 36172 handwritten data, and achieved 94.24% accuracy.	Pal, U. et al. (2007) [13]
5.	Identify the Kannada letters in the text by extracting it from the scanned images, then either store or display it for later use.	86% accuracy with Tesseract tool and 87% accuracy with Convolution Neural Network	Fernandes, R. et.al.(2019) [14]
6.	Using features based on modified log-Gabor filters applied at multiple scales and orientations, distinct scripts based on Matra and non-Matra are identified.	Positive findings demonstrate the effectiveness of the current tree-based method for classifying handwritten Indian scripts for each of the 12 officially recognized Indian Scripts.	Singh, P. K. et al. (2018) [15]
7.	5-layer CNN for a moderately large class character recognition problem	Accurate recognition performance is on par with state-of-the-art.	Maitra, D. S. et al. (2015) [16]
8.	preparing the obtained image and translating it into Kannada characters.	Gives the ability to modify Tulu archives through machine mapping of Kannada Unicode structure.	Antony, P. J. et al. (2016)[17]
9.	outlines several methods for handwritten recognition.	aids in understanding the characteristics of the ancient script and provides suggestions for future study	Antony, P. J. et al. (2016)[18]
10.	Digitizing tulu scripts using shallow and deep learning techniques	comparative study shows deep CNN gives high efficiency of conversion	Savitha, C.K. et al. (2018)[19]
11.	offers a technique for creating datasets.	incorporates manually written Tulu characters into models for machine learning	Sudarshan K et al.(2024) [20]
12.	English language converted to Tulu language	89% accuracy achieved in word based analysis and 81% accuracy in sentence based analysis	Rodrigues et al. (2024) [21]

V. METHODS FOR DIGITIZING HANDWRITTEN SCRIPTS

Digitizing handwritten scripts involves converting physical documents or texts written by hand into digital formats that can be stored, edited, and analyzed electronically. Here are some methods commonly used for digitizing handwritten scripts:

1. Scanning: Scanning handwritten documents is one of the most straightforward methods of digitization. Specialized document scanners or flatbed scanners can be used to capture images of handwritten pages. High-resolution scanning is preferred to capture finer details, especially for scripts with intricate characters or styles.
2. Image Processing and Optical Character Recognition (OCR): After scanning, image processing techniques can be applied to enhance the quality of captured images, such as adjusting brightness, contrast, and removing noise. Optical Character Recognition (OCR) software can then be employed to analyze the scanned images and convert them into editable and searchable text. While OCR technology has significantly advanced, its accuracy can vary depending on the handwriting style and quality of the original document.
3. Digitizing Tablets and Stylus Pens: For handwritten texts created digitally, such as handwritten notes on tablets or graphics tablets, the digital files are already in electronic format. These files can be directly stored, edited, and shared without the need for scanning and OCR. Digitizing tablets and stylus pens offer a natural way to write by hand while producing digital output in real-time.
4. Transcription by Hand: In cases where OCR technology may not be suitable due to the complexity of the script or poor handwriting quality, manual transcription by human operators becomes necessary. Skilled individuals proficient in the script can manually transcribe the handwritten text into digital formats using word processing software or specialized transcription tools.
5. Crowdsourcing: Crowdsourcing platforms can be utilized to engage a large number of contributors in digitizing handwritten scripts. Volunteers from the community or online platforms can transcribe handwritten texts into digital formats, contributing to the digitization process on a larger scale.
6. Handwriting Recognition Software: While traditional OCR focuses on printed text, handwriting recognition software is specifically designed to interpret handwritten characters. These software applications analyze handwriting strokes and patterns to recognize individual characters and words,

converting them into digital text. Training the software on specific handwriting styles can improve accuracy for particular scripts.

7. Combining Automated and Manual Methods: A combination of automated methods like OCR and manual transcription may offer the best results, especially for handwritten scripts with varying styles or historical documents with faded or deteriorated text. Automated methods can be used for initial digitization, while manual intervention can correct errors and ensure accuracy.
8. Annotation and Markup: In addition to transcribing handwritten texts into digital formats, annotation and markup tools can be used to add metadata, annotations, or additional information to the digitized documents. This enhances the usability and accessibility of the digitized materials for research, analysis, and archival purposes.
9. Each method for digitizing handwritten scripts has its advantages and limitations, and the choice of method depends on factors such as the nature of the handwritten texts, available resources, and desired level of accuracy and efficiency. Combining multiple methods may be necessary to achieve optimal results, particularly for challenging handwritten scripts.

VI. IMPORTANCE OF DIGITIZING TULU HANDWRITTEN SCRIPTS

1) The future of handwritten script digitization holds great promise, with ongoing research focusing on enhancing the accuracy of recognition systems and expanding the scope of digitization to include diverse linguistic scripts and styles. As the technology continues to evolve, it is poised to have an even greater impact on fields such as linguistics, , and cultural heritage preservation.
2) In conclusion, the digitization of Tulu handwritten scripts marks a pivotal moment in the preservation, promotion, and advancement of the Tulu language and culture. By converting traditional handwritten materials into digital formats, this process opens up a world of opportunities for the Tulu-speaking community and beyond.
3) First and foremost, digitization ensures the preservation of Tulu cultural heritage for future generations. Ancient manuscripts, literary works, and historical documents can now be safeguarded in digital archives, making them accessible to scholars, researchers, and enthusiasts worldwide. This preservation effort not only honors the rich linguistic tradition of Tulu but also provides invaluable insights into its history, customs, and folklore.
4) Furthermore, digitization fosters language revitalization and education initiatives within the Tulu-speaking community. Digital learning resources, interactive tools, and educational platforms in Tulu script empower learners of all ages to engage with their language in meaningful ways. As a result, fluency in Tulu can be cultivated, strengthening community identity and pride.
5) The digitization of Tulu handwritten scripts also facilitates scholarly research and linguistic analysis. Linguists and researchers can delve deeper into the structure and evolution of the Tulu language, uncovering unique linguistic features and cultural nuances. This academic exploration contributes to the broader field of linguistics while celebrating the distinctiveness of Tulu among the world's languages.
6) Moreover, digitization promotes digital inclusion and access to information for Tulu speakers worldwide. Through digital platforms and online communication channels, Tulu content can be shared, disseminated, and celebrated on a global scale. This global visibility not only raises awareness of the Tulu language and culture but also fosters cross-cultural dialogue and understanding.

VII. CASE STUDIES OF SUCCESSFUL TULU SCRIPT DIGITIZATION

There have been ongoing efforts to digitize various Indian scripts, including Tulu, as part of broader initiatives for language preservation and digital inclusion. Here are some potential case studies and initiatives that may provide insights into successful Tulu script digitization efforts:

1) Unicode Consortium: The Unicode Consortium is responsible for standardizing character encoding across different languages and scripts, including Tulu. Any successful digitization effort for Tulu script would likely involve encoding Tulu characters in the Unicode standard, ensuring interoperability across different platforms and systems.
2) Open Source Projects: There might be open-source projects focusing on Tulu script digitization, either as standalone efforts or as part of larger projects for digitizing Indian languages. These projects could provide valuable case studies on the challenges faced and solutions implemented in digitizing Tulu script.
3) Government Initiatives: Governments in India, particularly in Karnataka and Kerala where Tulu is spoken, often support initiatives for the preservation and promotion of local languages. Successful government-led projects related to Tulu language and script digitization could serve as case studies for understanding the institutional support and community engagement necessary for such efforts.
4) Academic Research: Universities and research institutions may undertake projects related to Tulu language and script digitization as part of linguistic research or language technology development. These projects could offer valuable insights into the linguistic characteristics of Tulu script and the technical challenges involved in digitizing it.

5) Community-Led Initiatives: Tulu-speaking communities themselves may initiate digitization efforts to preserve their language and cultural heritage. Community-led projects often leverage local knowledge and resources, fostering greater ownership and sustainability in the digitization process.
6) Digital Archives and Libraries: Digital archives and libraries focusing on South Indian languages may include collections of Tulu texts and resources that have been digitized. These repositories could provide examples of successful Tulu script digitization and offer access to digitized Tulu materials for research and education purposes.

While specific case studies of successful Tulu script digitization may be limited, examining broader initiatives in Indian language digitization and language preservation efforts can provide valuable insights and best practices applicable to Tulu script digitization. Additionally, staying updated with recent developments in the field of language technology and digital humanities may uncover new case studies and initiatives related to Tulu script digitization.

VIII. IDEAL SOLUTION, DESIRED STATUS & IMPROVEMENTS REQUIRED:

Digitizing Tulu handwritten scripts involves a comprehensive approach that combines advanced image processing techniques with state-of-the-art machine learning algorithms tailored to the intricacies of the Tulu script. This solution would yield accurate transcription of handwritten Tulu characters, ensuring faithful representation of the language in digital format. Additionally, the ideal solution would incorporate robust post-processing methods to refine the digitized text and enhance its usability for various applications.

Desired Status: The desired status for Tulu handwritten script digitization is a well-established ecosystem comprising efficient tools, standardized methodologies, and accessible resources. In this state, researchers and practitioners would have access to reliable datasets, cutting-edge algorithms, and collaborative platforms to advance the digitization efforts. The desired status also entails widespread adoption of digitized Tulu materials in educational, cultural, and scholarly contexts, contributing to language preservation and community empowerment.

Improvements Required:

1) **Enhanced Image Processing Techniques:** There is a need for improved image processing techniques specifically tailored to handle the unique characteristics of Tulu handwritten scripts. These techniques should effectively address challenges such as varying handwriting styles, script complexities, and image quality variations.
2) **Specialized Machine Learning Models:** Developing specialized machine learning models optimized for Tulu script recognition is essential. These models should leverage deep learning architectures and transfer learning approaches to achieve high accuracy in transcription tasks while accommodating the linguistic nuances of Tulu.
3) **Standardized Datasets:** The creation of standardized datasets comprising diverse samples of handwritten Tulu scripts is crucial for algorithm development and evaluation. These datasets should encompass various writing styles, historical documents, and contemporary texts to ensure the robustness and generalizability of digitization models.
4) **Community Engagement:** Active involvement of the Tulu-speaking community in digitization efforts is essential. Community engagement initiatives should promote awareness, participation, and collaboration in digitization projects, fostering a sense of ownership and cultural stewardship among stakeholders.
5) **Interdisciplinary Collaboration:** Collaboration between linguists, computer scientists, cultural heritage experts, and community members is vital for holistic digitization efforts. Interdisciplinary approaches can enrich digitization methodologies, address complex challenges, and ensure the cultural relevance and authenticity of digitized Tulu materials.
6) **Infrastructure and Resources:** Adequate infrastructure and resources, including funding, computing resources, and technical expertise, are necessary to support digitization initiatives effectively. Investments in infrastructure should prioritize accessibility, scalability, and sustainability to facilitate long-term digitization goals.
7) **User-friendly Tools and Interfaces:** Development of user-friendly tools and interfaces for digitizing, accessing, and utilizing Tulu handwritten scripts is essential. These tools should cater to diverse user needs, including researchers, educators, students, and cultural enthusiasts, while promoting ease of use and accessibility.
8) **Quality Assurance Mechanisms:** Implementation of quality assurance mechanisms to ensure the accuracy, reliability, and authenticity of digitized Tulu materials is critical. These mechanisms should incorporate validation protocols, error detection algorithms, and community feedback mechanisms to maintain data integrity and trustworthiness.
9) **Policy and Advocacy:** Advocacy for policy initiatives supporting Tulu language preservation and digitization is necessary at governmental, institutional, and organizational levels. Policies should promote language rights, cultural diversity, and digital inclusion, fostering an enabling environment for Tulu script

digitization initiatives.

10) **Long-term Sustainability:** Long-term sustainability strategies should be integrated into digitization plans to ensure the continued maintenance, updates, and accessibility of digitized Tulu materials. Sustainability efforts should encompass capacity building, knowledge sharing, and collaborative partnerships to safeguard the longevity of digitization endeavors.
11) **Cost performance ratio:** The work needs to consider the cost performance ratio and should identify the threshold values in order to overcome the Thrashing effect (Bhat, S., et al. (2021) [22]).

IX. RESEARCH GAP:

The current state of research in Tulu handwritten script digitization reveals several notable gaps:

1) Existing research lacks comprehensive datasets specifically tailored for Tulu handwritten script recognition, hindering the development and evaluation of digitization models.
2) There is a need for algorithmic adaptation to address the linguistic complexities and script variations inherent in Tulu handwriting, which may differ significantly from other Indic scripts.
3) Research often overlooks the cultural and linguistic context surrounding Tulu script usage, resulting in digitization models that may not adequately capture the nuances of Tulu handwriting.
4) The level of community engagement in digitization efforts is relatively low, indicating a gap in involving stakeholders in the co-creation and validation of digitization solutions.
5) The absence of standardized guidelines and protocols for Tulu script digitization poses challenges in ensuring consistency, interoperability, and scalability across digitization projects.
6) There is a lack of interdisciplinary collaboration between linguists, computer scientists, cultural experts, and community members, limiting the holistic understanding and approach to Tulu handwritten script digitization.
7) Clear evaluation metrics tailored to Tulu script digitization are lacking, making it difficult to assess the performance and effectiveness of digitization models accurately.

X. RESEARCH AGENDAS BASED ON RESEARCH GAP:

Based on the identified research gap, the following research agendas are proposed:

a. Initiate efforts to create comprehensive datasets of handwritten Tulu scripts, encompassing diverse writing styles, historical documents, and contemporary texts.
b. Develop novel machine learning algorithms and deep learning architectures optimized for Tulu script recognition, considering the linguistic characteristics and script variations specific to Tulu.
c. Incorporate cultural and linguistic context into digitization models and methodologies to ensure culturally sensitive and contextually relevant digitization solutions.
d. Implement community engagement strategies to involve Tulu-speaking communities in all stages of digitization projects, from dataset creation to model development and evaluation.
e. Establish standardized guidelines and protocols for Tulu script digitization, including data annotation, preprocessing techniques, and evaluation metrics, to ensure consistency and interoperability across projects.
f. Foster interdisciplinary collaboration between linguists, computer scientists, cultural experts, and community members to facilitate holistic approaches to Tulu handwritten script digitization.
g. Develop robust evaluation frameworks and benchmarks tailored to Tulu script digitization, incorporating linguistic accuracy, cultural relevance, and community feedback as key evaluation criteria.

XI. ANALYSIS OF RESEARCH AGENDAS:

The proposed research agendas align closely with the identified research gap and address key challenges in Tulu handwritten script digitization:

a. Comprehensive datasets are foundational for advancing digitization research and ensuring the robustness and generalizability of digitization models.
b. Tailoring machine learning algorithms to the linguistic nuances of Tulu script is essential for achieving accurate and contextually relevant digitization results.
c. Integrating cultural context into digitization methodologies fosters a deeper understanding of Tulu language and culture, enhancing the authenticity and relevance of digitized materials.
d. Engaging Tulu-speaking communities in digitization projects promotes community ownership, cultural stewardship, and sustainable digitization outcomes.
e. Standardized guidelines and protocols facilitate consistency, interoperability, and scalability in Tulu script digitization efforts, streamlining the digitization process and ensuring quality outcomes.
f. Collaborative efforts between diverse stakeholders enrich digitization research, combining linguistic expertise, technical proficiency, and cultural insights to address complex digitization challenges effectively.

g. Robust evaluation frameworks provide meaningful metrics for assessing the performance and impact of digitization models, guiding iterative improvements, and ensuring alignment with community needs and expectations.

By pursuing these research agendas, we can advance in the field of Tulu handwritten script digitization, contributing to language preservation efforts, and promote cultural heritage conservation within the Tulu-speaking community and beyond.

XII. SWOT ANALYSIS ON DIGITIZATION OF TULU SCRIPTS:

Strengths:

1. Digitization ensures the preservation of Tulu handwritten materials, protecting them from deterioration and loss.
2. Digital formats make Tulu handwritten scripts accessible to a wider audience, including scholars, researchers, and community members worldwide.
3. Digital texts enable efficient searchability and indexing, facilitating research and academic inquiries into Tulu language and culture.
4. Digitization supports language revitalization efforts by providing digital resources for Tulu language learning and literacy programs.
5. Digital platforms promote Tulu cultural heritage through digitized manuscripts, folklore, and artistic expressions, raising awareness and appreciation.
6. Digitization encourages community engagement in preserving and promoting Tulu language and culture, fostering a sense of cultural pride and identity.

Weaknesses:

1. Digitizing handwritten scripts may pose linguistic challenges, particularly in accurately transcribing and interpreting handwritten characters or scripts with complex writing systems.
2. Tulu script lacks a standardized script. While there have been efforts to standardize the script, it hasn't achieved widespread acceptance. This lack of standardization can make digitization efforts confusing and inconsistent.
3. Tulu script may not have readily available Unicode characters for all its letters and characters. Encoding these characters accurately and ensuring compatibility across different systems and platforms can be a significant challenge.
4. Even if Unicode characters exist for Tulu script, font support may be limited. Not all devices and software platforms may support Tulu fonts, making it difficult to display or render Tulu text correctly.
5. Developing efficient input methods for typing in Tulu script can be challenging. Users need convenient ways to input Tulu characters, especially considering the unique characters and diacritics present in the script.
6. Building a comprehensive digital corpus of Tulu texts is crucial for language preservation and development. However, collecting, and digitizing Tulu texts, especially older or rare texts, can be challenging due to factors such as availability, accessibility, and preservation issues.
7. Digitization efforts require resources such as funding, expertise, and infrastructure. Limited resources, especially in the context of a lesser-known language like Tulu, can pose significant barriers to effective digitization.

Opportunities:

1. Digitization opens opportunities for scholarly research and academic collaboration in Tulu studies, enhancing understanding and knowledge of Tulu language and culture.
2. Digitized Tulu handwritten scripts can serve as valuable educational resources for language learning, curriculum development, and cultural exchange programs.
3. Digital resources in Tulu script can support language revitalization efforts by making learning materials, textbooks, and educational content more accessible. Digital platforms can facilitate Tulu language instruction, literacy programs, and language documentation projects, fostering greater proficiency and usage of the language.
4. Digital archives of Tulu handwritten scripts enable researchers, linguists, and scholars to conduct in-depth studies on Tulu language, culture, and history. Digital tools and resources facilitate linguistic analysis, comparative studies, and corpus linguistics research, contributing to a deeper understanding of Tulu and its place within the broader linguistic landscape.
5. The digitization of Tulu handwritten scripts empowers the Tulu-speaking community to actively participate in preserving and promoting their language and cultural heritage. Community members can contribute to digitization efforts by transcribing texts, creating digital resources, and sharing knowledge through online platforms and social media.
6. Digital platforms and social media offer new avenues for Tulu speakers to communicate and share

content in their language. Digital tools and applications supporting Tulu script enable users to communicate through messaging, social networking, and digital publishing, fostering a vibrant online presence for Tulu language and culture.

7. Digital access to Tulu content facilitates communication, information exchange, and access to services for Tulu speakers, particularly those residing in geographically dispersed or diaspora communities. Digital platforms can provide news, government services, and online resources in Tulu script, promoting linguistic inclusion and digital equity.
8. The digitization of Tulu handwritten scripts drives technological innovation and development in language technology, optical character recognition (OCR), machine translation, and natural language processing (NLP). Advances in these areas benefit not only Tulu but also other lesser-known languages facing similar digitization challenges.
9. Digitized Tulu handwritten scripts contribute to the visibility and recognition of the Tulu language and culture on a global scale. Digital platforms and repositories make Tulu literature, poetry, and artistic works accessible to a broader audience, fostering cross-cultural exchange and appreciation.

Threats:

1. Disparities in digital access and literacy may exacerbate inequalities in accessing digitized Tulu handwritten scripts, marginalizing certain segments of the Tulu-speaking community.
2. Rapid advancements in technology may lead to the obsolescence of digital formats or platforms, potentially jeopardizing the long-term accessibility and preservation of digitized Tulu handwritten materials.

While the digitization of Tulu handwritten scripts offers significant opportunities for preservation, accessibility, and cultural promotion, it also poses challenges related to technology, linguistics, and digital inclusion. Addressing these challenges effectively requires collaborative efforts from stakeholders to ensure the successful digitization and sustainable preservation of Tulu language and culture.

XIII. SUGGESTIONS TO IMPLEMENT RESEARCH ACTIVITIES:

(1) Dataset creation
(2) Develop algorithms
(3) Engage people in research activities

XIV. CONCLUSION:

In essence, the digitization of Tulu handwritten scripts signifies a transformative shift towards a more interconnected, inclusive, and digitally empowered future. It represents a commitment to preserving cultural heritage, promoting linguistic diversity, and harnessing the power of technology for the greater good. As we continue this digitization journey, let us embrace the opportunities it brings and work together to ensure a vibrant and flourishing future for the Tulu language and its community.

The digitization of Tulu handwritten scripts has far-reaching implications for language preservation, cultural heritage, education, research, community empowerment, and technological advancement. By leveraging digital technologies and collaborative efforts, the Tulu-speaking community can chart a path towards a more inclusive, interconnected, and digitally empowered future for their language and culture.

REFERENCES:

[1] Xiao, G., Kuan, T. W., Li, F., Ling, M., & Tseng, S. P. (2022, November). Systematic Review of Machine Learning Methods in Customer Satisfaction Prediction. In *2022 10th International Conference on Orange Technology (ICOT)* (1-3). GoogleScholar↗

[2] Schaeffer, S. E., & Sanchez, S. V. R. (2020). Forecasting client retention—A machine-learning approach. *Journal of Retailing and Consumer Services*, *52*(1), 2-24. GoogleScholar↗

[3] Sharma, R., & Kaushik, B. (2020). Offline recognition of handwritten Indic scripts: A state-of-the-art survey and future perspectives. *Computer Science Review*, *38*, 100302. GoogleScholar↗

[4] Bhat, S. S., Ananth, A., Nambiar, R., & Bhat, N. (2022). Building dataset and deep learning-based inception model for the character classification of tigalari script. In *Recent Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE 2020* (pp. 239-252). Springer Singapore. GoogleScholar↗

[5] Sudarsan, D., Vijayakumar, P., Biju, S., Sanu, S., & Shivadas, S. K. (2018, March). Digitalization of malayalam palmleaf manuscripts based on contrast-based adaptive binarization and convolutional neural networks. In *2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 1-4). IEEE.GoogleScholar↗

[6] Babu, N., & Soumya, A. (2019, April). Character recognition in historical handwritten documents–a survey. In *2019 international conference on communication and signal processing (ICCSP)* (pp. 0299-0304). IEEE. GoogleScholar↗

[7] Bhat, S. S., Ananth, A., Nambiar, R., & Bhat, N. (2022). Building dataset and deep learning-based inception model for the character classification of tigalari script. In *Recent Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE 2020* (pp. 239-252). Springer Singapore. GoogleScholar↗

[8] Kataria, B., & Jethva, H. B. (2018). Review of Advances in Digital Recognition of Indian Language Manuscripts. GoogleScholar↗

[9] Tajne, N., & Mahajan, K. (2020). A Systemic Study of Indian Ancient Script by determining the Issues for Effective use of Handwritten Character Recognition. Niket Kirti Mahajan, 1–8. GoogleScholar↗

[10] Sinwar, D., Dhaka, V. S., Pradhan, N., & Pandey, S. (2021). Offline script recognition from handwritten and printed multilingual documents: a survey. *International Journal on Document Analysis and Recognition (IJDAR)*, *24*, 97-121.GoogleScholar↗

[11] Agnihotri, V. P. (2012). Offline handwritten Devanagari script recognition. *IJ Information Technology and Computer Science*, *8*(1), 37-42.GoogleScholar↗

[12] Singh, S., Sharma, A., & Chauhan, V. K. (2023). Indic script family and its offline handwriting recognition for characters/digits and words: a comprehensive survey. *Artificial Intelligence Review*, *56*(Suppl 3), 3003-3055. GoogleScholar↗

[13] Pal, U., Sharma, N., Wakabayashi, T., & Kimura, F. (2007, September). Off-line handwritten character recognition of devnagari script. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)* (Vol. 1, pp. 496-500). IEEE. GoogleScholar↗

[14] Fernandes, R., & Rodrigues, A. P. (2019, August). Kannada handwritten script recognition using machine learning techniques. In *2019 IEEE international conference on distributed computing, VLSI, electrical circuits and robotics (DISCOVER)* (pp. 1-6). IEEE. GoogleScholar↗

[15] Singh, P. K., Sarkar, R., Bhateja, V., & Nasipuri, M. (2018). A comprehensive handwritten Indic script recognition system: a tree-based approach. *Journal of Ambient Intelligence and Humanized Computing*, 1-18. GoogleScholar↗

[16] Maitra, D. S., Bhattacharya, U., & Parui, S. K. (2015, August). CNN based common approach to handwritten character recognition of multiple scripts. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 1021-1025). IEEE. GoogleScholar↗

[17] Antony, P. J., Savitha, C. K., & Ujwal, U. J. (2016, May). Haar features based handwritten character recognition system for Tulu script. In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 65-68). IEEE. GoogleScholar↗

[18] Antony, P. J., & Savitha, C. K. (2016, June). A framework for recognition of handwritten South Dravidian Tulu script. In *2016 Conference on Advances in Signal Processing (CASP)* (pp. 7-12). IEEE. GoogleScholar↗

[19] Antony, P. J., Savitha, C. K., & Ujwal, U. J. (2018). Efficient binarization technique for handwritten archive of South Dravidian Tulu script. In *Emerging Research in Computing, Information, Communication and Applications: ERCICA 2016* (pp. 651-666). Springer Singapore. GoogleScholar↗

[20] Sudarshan, K., & Bhat, S. (2024). A Review on Leveraging Handwritten Tulu Characters for Machine Learning in Optical Character Recognition. *Journal homepage: https://gjrpublication. com/gjrecs*, *4*(01). GoogleScholar↗

[21] Rodrigues, A. P., Vijaya, P., & Fernandes, R. (2024). Tulu Language Text Recognition and Translation. IEEE Access

[22] Bhat, S., & Vijayakumar, V. (2021, January 28). Adaptive Directory based Cache Coherency Model to Optimize the Network Bandwidth in Distributed Applications. *IOSR Journal of Engineering (IOSRJEN)*, 8(9), 41-46. https://ssrn.com/abstract=3774752 GoogleScholar↗

****