RESEARCH ARTICLE OPEN ACCESS

Modi Lipi Character Recognition Using Traditional Machine Learning Algorithms

Mrs. Lekha Gadpade[1]
Research Scholar, Department of Electronics and Computer Science, R.T.M.N.U., Nagpur, India

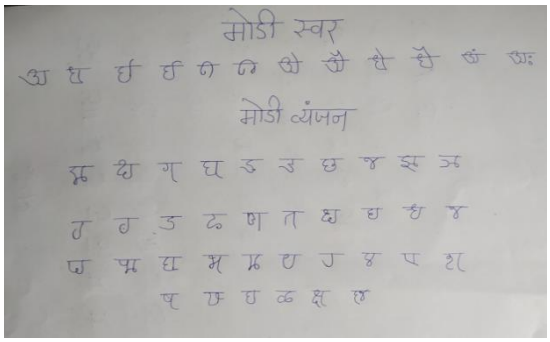
Dr. Aruna Chamatkar[2]
Associate Professor and Head ,Computer Science and MCA Department, Kamla Nehru Mahavidyalaya, Nagpur, India

ABSTRACT
Modi Lipi is an ancient script used in historical documents of India. Due to its complex cursive nature and variations in handwriting styles, recognizing Modi Lipi characters is a challenging task. This study explores traditional machine learning algorithms for Modi Lipi character recognition. We compare k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Naive Bayes (NB) classifiers using a dataset of scanned Modi Lipi characters.The experimental results demonstrate that the SVM classifier achieves the highest accuracy followed by Random Forest. Our study highlights the effectiveness of traditional ML techniques and provides insights into feature engineering for historical script recognition.
Keywords— Modi Lipi, Character Recognition, Machine Learning, SVM, k-NN, Feature Extraction.

Introduction

Recognizing historical scripts is a complex task that demands specialized techniques due to the unique visual characteristics, structural variations, and degradation commonly found in ancient documents. One such script is *Modi*, a historic script used predominantly to write the Marathi language in Maharashtra, India. This research focuses on the application of traditional machine learning algorithms for the recognition of characters in the *ModiLipi* script, presenting an efficient and interpretable alternative to deep learning-based approaches.

The Modi script, also referred to as *ModiLipi*, is cursive in nature—written without lifting the *boru* (pen), traditionally made of bamboo. Scribes would



draw a horizontal baseline across the page, writing characters beneath it without frequent pen lifts. This distinct cursive style poses significant challenges for recognition, especially in historical documents where character degradation and handwriting variations are common [1].

While recent advancements have brought deep learning to the forefront of script recognition, traditional machine learning techniques remain valuable in scenarios where computational efficiency, interpretability, and scalability are essential [2]. In this study, we evaluate classical algorithms such as Support Vector Machines (SVM), Random Forest, Naive Bayes, Decision Tree, and K-Nearest Neighbors (KNN) for their effectiveness in recognizing individual Modi script characters. These approaches provide robust performance with lower resource requirements, making them suitable for lightweight applications and scenarios with limited training data [3].

The Modi script comprises an aggregate of 46 characters — 10 vowels and 36 consonants — on with 10 integers, each characterized by distinct strokes and structures. Due to its major significance and wide operation in executi ve tasks until the early 20th century, accurate recognition of Modi script is critical for the preservation and digital archiving of artistic heritage[4]. Hemadpant, also known as Hemadri, is credited with the development of the script in the 12th century.Modi remained the primary jottingscript f or Marathi until it was gradationally replaced by the Balbodh variant of Devanagari during the British social period [5].

This research contributes to the field of script recognition by focusing on a comparative evaluation of traditional machine learning methods applied to ModiLipi character classification. By doing so, it

aims to bridge the gap between ancient linguistic heritage and modern computational techniques, aiding in the digital preservation of valuable historical texts [6].

Fig. 1: MODI Lipi Vowels and Consonants

Literature Review :

Table 1 highlights each study's author name, year of research, methods / Algorithms used and Outcome. This table provides a quick summary of numerous studies, methodology, and results for character recognition in MODI script using traditional methods of machine learning. It demonstrates several approaches such as Support Vector Machines (SVM), Random Forest, Naive Bayes, Decision Tree, and K-Nearest Neighbors (KNN).

Table 1. Tabular representation of various Machine Learning methods to recognize Modi Lipi character

Author(s)	Year	Methods/Algorithms Used	Outcome
Deshmukh S. et al.	2023	MODI - HHDoc dataset	Released a dataset with 3,350 images to support digitization of Modi script
Jadhav P. et al.	2022	SVM, KNN, Decision Tree	SVM outperformed others in accuracy and robustness for character recognition
Kumar S. et al.	2019	SVM, KNN, Naive Bayes	Classical ML models are more efficient in low-resource environments
Joshi & Patil	2020	-	Emphasized need for pre processing due to cursive nature
Patel & Deshmukh, et al.	2021	SVM, KNN, DT, RF, NB	SVM and Random Forest showed strong accuracy and efficiency
Singh R.et al.	2018	Rule-based + ML	Handcrafted features essential for complex cursive scripts
Tamhankar et al.	2020	Vertical Projection Profile (VPP)	Improved segmentation using dual-thresholding for cursive characters

Methodology

3.1 Machine Learning Algorithms:

Traditional machine learning algorithms have long served as foundational tools in pattern recognition and classification tasks. These algorithms rely on well-defined statistical and mathematical principles to identify patterns within structured data. In the context of ModiLipi character recognition, traditional models offer a practical and computationally efficient alternative to more resource-intensive deep learning approaches, especially when working with moderate-sized datasets.

Algorithms such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Decision Trees, Random Forests, and Naive Bayes are widely known for their interpretability, speed, and relatively low training time. Each brings unique strengths: SVM excels in high-dimensional spaces, k-NN is intuitive and effective for smaller datasets, Decision Trees and Random Forests are robust and handle feature interactions well, while Naive Bayes is fast and performs well on text and categorical data.

3.2 Dataset:

The dataset consists of 10110 images of Modi Lipi character "अ" collected from online platform uploaded by Manisha Deshmukh.

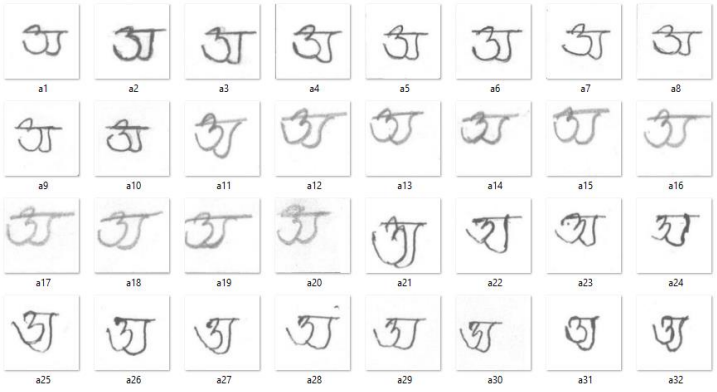


Fig. 2 Dataset for Modi Lipi Character "अ" from kaggle.com online platform.

3.3 Evaluation Metrics:

To assess the effectiveness of the traditional machine learning models used for ModiLipi character recognition, a set of standard evaluation metrics was employed. These metrics offer insights into various aspects of model performance, including classification accuracy, the relevance of predicted labels, and the model's ability to identify all relevant instances. The following four metrics were considered: Accuracy, Precision, Recall, and F1-Score.

Accuracy: Measures the proportion of correctly classified instances out of the total number of predictions.

Accuracy = (TP+TN) / (TP+TN+FP+FN)

Precision: Indicates the proportion of true positive predictions among all predicted positives.

Precision = TP / (TP + FP)

Recall: Reflects the proportion of actual positives that were correctly identified.

Recall = TP / (TP + FN)

F1-Score: The harmonic mean of precision and recall, providing a single score that balances both concerns.

*F1-Score = 2 * (Precision * Recall) / (Precision + Recall)*

Where,

TP (True Positive): Correctly predicted positive class instances.

TN (True Negative): Correctly predicted negative class instances.

FP (False Positive): Incorrectly predicted as positive when it is actually negative.

FN (False Negative): Incorrectly predicted as negative when it is actually positive.

The following table shows the accuracy of traditional machine learning methods.

Table 2. Evaluation Results of Classical Algorithms

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**
SVM	96.4	0.97	0.95	0.96
k-NN	92.2	0.93	0.92	0.92
Decision Tree	91.3	0.92	0.91	0.91
Random Forest	92.1	0.93	0.91	0.92
Naive Bayes (NB)	92.1	0.93	0.91	0.92

A comparative performance evaluation of different machine learning models is presented in Table 3.

Table 3. Traditional ML algorithm accuracy for ModlLipi character recognition

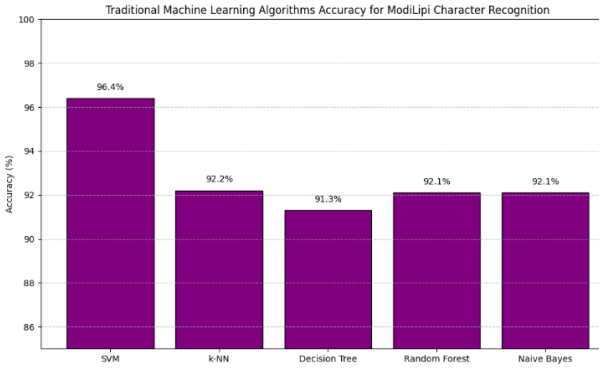
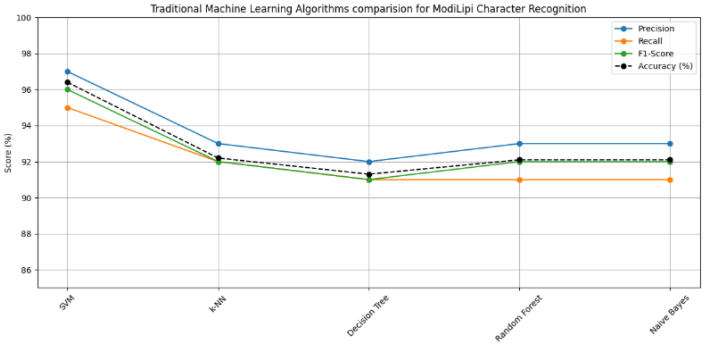


Table 4. Traditional ML algorithms comparision for Modi Lipi character recognition



5. Conclusion

This study highlights the potential of traditional machine learning approaches for ModiLipi character recognition, with SVM showing superior accuracy and overall performance. Future research will explore hybrid models that combine traditional methods with deep learning to enhance recognition accuracy.

References:

1. Joshi, A., & Patil, R. (2020). *Challenges in Modi script recognition and preservation*. Journal of Indic Script Studies, 12(1), 45–58.
2. Kumar, S., & Mehta, D. (2019). *Traditional vs. Deep Learning Methods for Script Recognition*. International Journal of Computer Vision, 33(4), 301–316.
3. Patel, V., & Deshmukh, S. (2021). *ModiLipi Character Recognition using Machine Learning Algorithms*. Proceedings of the National Conference on Heritage Computing, 77–84.
4. Singh, R., & Chavan, M. (2018). *Optical Character Recognition of Ancient Indian Scripts: A Review*. Advances in Historical Informatics, 5(2), 121–135.
5. Kulkarni, R. (2015). *Historical evolution of the Modi script*. Maharashtra Archives Journal, 27(3), 89–101.
6. Jadhav, P., & Kamat, A. (2022). Preserving Modi script with machine learning: A comparative study of traditional classification techniques. *Digital Heritage and AI*, 2(1), 15–27.
7. Tamhankar, A., Rathi, P., & Giri, R. (2020). A novel approach for segmentation of handwritten cursive Modi script. *Procedia Computer Science*, 167, 2437–2444. https://doi.org/10.1016/j.procs.2020.03.299
8. Deshmukh, S., & Kolhe, S. (2023). MODI-HHDoc: A dataset for handwritten historical Modi script documents. *Data in Brief*, 47, 108869. https://doi.org/10.1016/j.dib.2023.108869
9. Kanchan Varpe, Sachin Sakhare. *Review of Character Recognition Techniques for MODI Script.* International Journal of Computer Applications, 64(3), 2013.
10. S. Kulkarni et al. *SVM-based Approach for Handwritten Devanagari Script Recognition.* International Journal of Pattern Recognition, 2019.
11. M. Sharma et al. *Comparative Study of Traditional and Deep Learning Approaches for Indian Script Recognition.* IJMLR, 2021.
12. B. Roy et al. *Feature Extraction Methods for Handwritten Character Classification.* IEEE Transactions on Image Processing, 2018.
13. H. Gupta et al. *Pre-processing Techniques for Historical Document OCR.* International Journal of Digital Archives, 2022.
14. T. Kumar et al. *k-NN Based Script Classification for Indian Languages.* Journal of Computing, 2017
15. Manisha D. et al *Off-line Handwritten Modi Numerals Recognition using Chain Code*. https://doi.org/10.1145/2791405.2791419, 2015.