International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 04 | Apr 2023 www.irjet.net p-ISSN: 2395-0072

Survey On Broken and Joint Devanagari Handwritten Characters Recognition Using Deep Learning

Prachi Pachang[1], Jiya Shaikh[2], Ms. Vina M. Lomate[3], Tanishka Sinha[4], Manjeet Kour[5]

[3]Hod Dept. of Computer Engineering, RMD Sinhgad School of Engineering, Warje Pune, India
[1,2,4,5]UG Student, Computer Engineering, RMD Sinhgad School of Engineering, Warje Pune, India

---

Abstract - *The recognition of handwritten Devanagari characters presents a significant challenge due to the script's complexity and variability. The complexity is further compounded by the variability of broken and joint characters that are written differently by different individuals. In recent years, deep learning models have emerged as a powerful solution for character recognition, achieving remarkable performance in various applications. This survey paper presents an in-depth analysis of the deep learning-based approaches used for recognizing handwritten Devanagari broken and joint characters. We extensively review the architectures and techniques applied in deep learning models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models, to identify these characters. We also discuss the datasets utilized for training and testing these models and the performance metrics used for evaluating their performance. Additionally, we conduct a comparative analysis of the different approaches, highlighting their respective strengths, and limitations, and proposing possible directions for future research. Our survey is intended to serve as a valuable resource for researchers and practitioners engaged in the area of handwritten Devanagari character recognition using deep learning.*

Key Words: **Feature extraction, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), TensorFlow, ImageDataGenerator, Text recognition**, **Wavelets**.

1. INTRODUCTION

Handwritten character recognition is vital in computer vision and pattern recognition, with practical applications such as optical character recognition, automatic form processing, and intelligent handwriting recognition systems. Devanagari is a prominent script used in several languages such as Hindi, Marathi, and Nepali, and recognizing handwritten Devanagari characters is a challenging task due to the complexity and variability of the script. The recognition of broken and joint characters in Devanagari further adds to the complexity, as these characters are often written differently by different individuals, making it difficult to develop a robust recognition system.

This survey paper aims to provide a comprehensive overview of the existing approaches for broken and joint handwritten Devanagari character recognition, with a particular focus on the involvement of wavelet transform and recent deep learning-based technologies. The paper will also discuss the advantages and limitations of each approach and highlight the current techniques. Additionally, publicly available datasets such as the Devanagari Handwritten Character Dataset (DHCD) and Indian Language Handwritten Character Dataset (ILHCD) will be reviewed, which have been widely used for training and evaluation of various approaches. This survey paper will provide a useful resource for researchers and practitioners working in the field of handwritten Devanagari character recognition, with the aim of improving the accuracy and efficiency of character recognition systems.

2. Main Terminologies:

2.1 Devanagari script: A script used for writing several languages, including Hindi, Marathi, and Nepali.

2.2 Broken characters: Devanagari characters that are separated or disjointed, which require additional techniques for recognition.

2.3 Joint characters: Devanagari characters that are connected to other characters, often written in a cursive manner, require additional techniques for recognition.

2.4 Handwritten character recognition: The process of identifying and transcribing handwritten characters from an image or document.

2.5 Devanagari Handwritten Character Dataset (DHCD): A publicly available dataset containing handwritten Devanagari characters.

3. Literature Survey:

Sr. no	Publication details	Tech used	Dataset	Accuracy	Research Gap Identified
1	Performance Evaluation of Learning-based Frameworks for Devanagari Character Recognition Saptarshi Kattyayan, P. Kanungo	Prepossessing: CNN Model, De-noising, Size, and Contrast Retuning	The Devanagari Character Dataset includes characters from three distinct classes: Vowels, Consonants, and Numerals. The numerical Dataset consists of 10 classes ranging from 0 to 9 for each digit there are 2000 samples present in the dataset. Vowels dataset consists of 12 classes containing 2000 samples in a given class. The consonant dataset contains the highest data. Each Consonant contains 2000 samples.	98.01%	There is a need for an efficient character recognition and classification system.
2	Character Recognition System for Devanagari Script Using Machine Learning Approach Shilpa Mangesh Pande, Bineet Kumar Jha	Pre-Processing: Normalization, Thinning and noise removal. Classification: Decision tree classifier, Nearest Centroid classifier, K Neighbors Classifier, Extra tree Classifier	There is a scanned Devanagari script alphabets database consisting of 43 thousand of images of 32x32 pixels	78%	For complex model deep learning can be used .
3	Deformed character recognition using convolutional neural networks	Pre-Processing: Data Augmentation Classification: Tree Classifier, SVM Classifier	The datasets employed for training in case of printed data samples are extracted from ancient Kannada documents whereas the handwritten data sample are collected in varied environments	98.05%	Only 52 classes are present, which does not represent fully complexity of the recognition problem

4	Handwritten Devanagari Character Recognition using Wavelet-Based Feature Extraction and Classification Scheme Adwait Dixit, Ashwini Navghane, Yogesh Dandawate IEEE India Conference (INDICON)	Prepossessing: Banalization Feature Extraction: Wavelet Transform Classification: ANN, OCR	There is a dataset, which contains almost 2000 different characters taken from different people for each 20 characters	70%	Only Devanagari characters were considered. Further research can be on other Indian regional languages
5	Transfer Learning using CNN for Handwritten Devanagari Character Recognition	Feature Extraction: AlexNet, DenseNet, Vgg, Inception, ConvNet	The dataset of Devanagari has 46 classes. Each class has 2000 images. The dataset consists of 92000 images. 78200 images for training and 13800 for testing.	98%	Possibility of overfitting and limited sets of pre trained model.
6	Recognition of Handwritten Characters Based on Wavelet Transform and SVM Classifier, The International Arab Journal of Information Technology, Vol. 15, No. 6, Malika Ait Aider, Kamal Hammouche, and Djamel Gaceb	Feature extraction: Wavelet transform Pre-Processing: Normalization Classification: SVM	MNIST Dataset was used	98%	The paper suggests future work on integrating a normalization operation as a preprocessing procedure, but this was not explored in the current study.
7	Handwritten Devanagari Character Recognition Using Layer-Wise Training of Deep Convolutional Neural Networks and Adaptive Gradient Methods	Pre-processing involves using deep convolutional neural networks. Feature extraction is done through the box approach, dividing the character into 24 cells.A normalized vector distance is computed for each box, except the empty cells.	ISIDCHAR: a database with 36,172 grayscale images of 47 Devanagari characters. V2DMDC: a database with 20,305 samples of handwritten Devanagari characters.	98%	Devanagari characters only were used.

8	Marathi Handwritten Character Recognition Using SVM and KNN Classifier. Diptee Chikmurge; R. Shriram. Springer HIS Advances in Intelligent Systems and Computing, vol 1179; Published	K-Nearest Neighbours and SVM	The dataset of Marathi handwritten characters available on Kaggle consists of 58,000 images of characters, covering a total of 58 different types of Marathi characters.	The accuracy achieved was 90% for KNN and 95% for SVM.	The drawback of HOG is its slow computation speed, which can be addressed by employing an alternative technology.
9	Handwritten Marathi Compound Character Recognition. Amol A. Kadam, Dr. Milind V. Bhalerao, Mohit N. Tanurkar, IJERT	Classifiers used were SVM and KNN.	There are 3500 images of compound Marathi characters written by hand.	The SVM classifier achieved an accuracy of 96.49%, while the KNN classifier achieved an accuracy of 95.67%.	The number of features is low.
10	Handwritten Marathi character (vowel) recognition; Ajmire P.E. and Warkhede S.E.	Gaussian Distribution Function	There are 120 images that show Marathi vowels in different styles.	60%	The accuracy is not satisfactory and requires further optimization.

4. Algorithmic Survey:

Sr no.	Publication Details	Algorithm Used	Accuracy	Language
1	Performance Evaluation of Learning-based Frameworks for Devanagari Character Recognition. Saptarshi Kattyayan, P. Kanungo	Gaussian Naive Bayes, Decision Tree, K-Nearest Neighbor(KNN) and CNN	98.01%	Devanagari
2	Character Recognition System for Devanagari Script Using Machine Learning Approach. Shilpa Mangesh Pande, Bineet Kumar Jha	Decision Tree classifier, Nearest Centroid classifier, K Nearest Neighbors classifier, Extra Trees classifiers and Random Forest classifier	78%	Various Scripts like Handwritten Devanagari , Arabic, English and Chinese
3	Deformed character recognition using convolutional neural networks			Kannada Dataset
4	Handwritten Devanagari Character Recognition using Wavelet-Based Feature Extraction and Classification Scheme Adwait Dixit, Ashwini Navghane, Yogesh Dandawate IEEE India Conference (INDICON)	ANN with Wavelet feature	70%	Handwritten Devanagari Script
5	Transfer Learning using CNN for Handwritten Devanagari Character Recognition Nagender Aneja and Sandhya Aneja Universiti Brunei Darussalam Brunei Darussalam	DCNN	98%	Handwritten Devanagari

6	Recognition of Handwritten Characters Based on Wavelet Transform and SVM Classifier, The International Arab Journal of Information Technology, Vol. 15, No. 6, Malika Ait Aider , Kamal Hammouche , and Djamel Gaceb	CWT (Continuous Wavelet Transform), SVM, K-nearest neighbor	98%	Off-line Handwritten Characters
7	Handwritten Devanagari Character Recognition Using Layer-Wise Training of Deep Convolutional Neural Networks and Adaptive Gradient Methods Mahesh Jangid, Sumit Srivastava	DCNN and Adaptive Gradient Methods	98%	Devanagari Script
8	Marathi Handwritten Character Recognition Using SVM and KNN Classifier. Diptee Chikmurge; R. Shriram. Springer HIS Advances in Intelligent Systems and Computing, vol 1179; Published in 2020	SVM (Support Vector Machine) and KNN (K - Nearest Neighbours)	The accuracy achieved was 90% for KNN and 95% for SVM.	Marathi

9	Handwritten Marathi Compound Character Recognition. Amol A. Kadam, Dr. Milind V. Bhalerao, Mohit N. Tanurkar, IJERT	Classifiers used were SVM and KNN.	The SVM classifier achieved an accuracy of 96.49%, while the KNN classifier achieved an accuracy of 95.67%.	Marathi
10	Handwritten Marathi character recognition using R-HOG Feature; Parshuram M. Kamble,Ravindra S. Hegadi	SVM and FFANN	The TAR calculated to be 97.15% for FFANN and 95.64% for SVM respectively.	Marathi

5. Live Survey:

The 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) by IEEE includes experimentation on broken and joint handwritten character recognition using deep learning. The experiment involves testing various CNN architectures with different depths and structures. The results are then compared with state-of-the-art methods like VGG16, VGG19, InceptionV3, MobileNet, ResNet50, and Xception using transfer learning of pre-trained weights.

6. Conclusion:

In conclusion, there has been significant research in the field of broken and joint handwritten character recognition using deep learning. Several studies have utilized various deep learning architectures, such as convolutional neural networks (CNNs), to improve the accuracy of recognition models. Techniques such as feature extraction, pre-processing, and transfer learning have also been employed to improve model performance. The results of these studies indicate that deep learning-based approaches can achieve high accuracy rates in recognizing broken and joint characters in handwritten scripts. However, more research is still needed to optimize the performance of these models and make them more efficient and reliable for real-world applications.

7. References:

1. S. Kattyayan, T. Kar and P. Kanungo, "Performance Evaluation of Learning Based Frameworks for Devanagari Character Recognition," 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2020

2. P. Gupta, S. Deshmukh, S. Pandey, K. Tonge, V. Urkunde and S. Kide, "Convolutional Neural Network based Handwritten Devanagari Character Recognition," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020, pp. 322-326

3. N. Aneja and S. Aneja, "Transfer Learning using CNN for Handwritten Devanagari Character Recognition," 2019 1st International Conference on Advances in Information Technology (ICAIT), 2019

4. R. Karnik, "Recognition of Handwritten Devanagari Characters", Fifth International Conference on Document Analysis and Recognition ICDAR '99

5. Rani, N Shobha & Chandan, Nagabasavanna & Jain, Sajan & Kiran, Hena. (2018). Deformed character recognition using convolutional neural networks. International Journal of Engineering & Technology. 7. 1599. 10.14419/ijet.v7i3.14053.

6. Recognition of broken and overlapping handwritten Bangla digits using convolutional neural network by S. R. Chowdhury and S. M. A. Bhuiyan, published in the International Journal of Advanced Computer Science and Applications (IJACSA), 2019.

7. "Overlapping handwritten character recognition using CNN with geometric normalization" by M. Khan and A. Aziz, published in the International Journal of Advanced Computer Science and Applications (IJACSA), 2018.

8. "Broken Bangla handwritten character recognition using CNN with data augmentation" by M. R. Rahman, M. A. Rahman, and M. Z. Rahman, published in the International Conference on Intelligent Systems Design and Applications (ISDA), 2020.

9. "Recognition of overlapping and broken handwritten digits using deep learning" by B. H. N. R. Ramasamy, M. S. T. Khan, and S. S. Ali, published in the International Conference on Signal Processing and Intelligent Systems (ICSPIS), 2019.

10. "A review on overlapping and touching character recognition using deep learning" by A. B. A. Azhar, M. N. M. Nasir, and N. A. M. Isa, published in the Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2020.

11. "A comprehensive survey on overlapping and broken character recognition" by A. Sharma and M. Vatsa, published in the Journal of Pattern Recognition Research, 2021.

12. "Recognizing Overlapping Handwritten Digits Using Convolutional Neural Networks" by L. Mai, X. Chen, and D. D. Feng, published in the IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2018.