International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1827-1830 www.ijaem.net ISSN: 2395-5252

Handwritten Marathi Character Recognition using SVM Classifier

Akanksha Kulkarni, Payal Giri, Tasmiya Pathan, Syed Tabassum, Amol Dhumane
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune

Date of Submission: 20-03-2023 Date of Acceptance: 30-03-2023

ABSTRACT:
The Optical Character Recognition (OCR) technique is used forrecognizing handwritten characters inany language like English, Chinese, Korean, etc. However, it has recently made significant advancements and is now much more difficult for languages like Marathi. Therefore, it is important to concentrate on languages of Indian origin. This paper'smain focus is the classification of handwritten Marathi characters using various methods. The procedure's execution involves a number of processes, including character segmentation, feature extraction, and classifications to identify characters with the appropriate styles. To determine which classifier is the most accurate, the comparison accuracy of SVM with other classifiers is investigated.

KEYWORDS
OCR (Optical Character Recognition), MODI Script, K-Means Clustering, FFANN (Feedforward Artificial Neural Network),R-HOG (Rectangle Histogram Oriented Gradient), SVM (Support Vector Machine),KNN (K-Nearest Neighbours).

I.INTRODUCTION

Character recognition is a widely known study field that has remarkable accuracy when comparing printed text with handwritten text. The use of scripts from outside India, such as Chinese, Japanese, Korean, German, etc., has already become more common, but it is still important to focus on characters used in Indian languages. Banks, post offices, libraries, and periodicals are just a few of the real-world and business uses for automatic character recognition of printed and handwritten information such as envelopes, checks, application forms, and others.

The OCR categories include offline and online character recognition, and offline character reading is further categorized into two groups: machine printed and handwritten. Character identification of handwritten papers has many issues compared to machine-printed texts because of the structure, shape, multiple strokes, and various writing styles. The Devanagari inscription is where the Marathi language gets its start. The oldest Devnagari script used to write several other languages, including Sindhi, Pali, Hindi, Konkani, and others. As Marathi is a widely spoken language in Maharashtra, the study of several classifiers employed in the recognition of automatically handwritten characters in Marathi is the main emphasis of this paper.

II. LITERATURE SURVEY

In paper[1],the model proposed by the authors forrecognizingthe Handwritten MODI Script by using CNN(Convolutional Neural Network) autoencoder as a feature extraction technique and SVM(Support Vector Machine) as a classifier. Until 1950, the Marathi language was written in MODI Script, avery old Indian script.CNN autoencoder is used for feature extraction of the input image, which has a size of 60*60 pixels. The data augmentation is thenperformedon the input image having a size of 60*60 pixels. After performing the augmentation method, the data is given to the CNN autoencoder for applying the feature extraction technique. The feature vectors extracted are given to SVM classifier with RBF kernel for recognizing the characters. The accuracy achieved is 99.3%

In paper [2], authors have recognized handwritten Hindi characters using K-Means Clustering as a Feature Extraction technique and SVM with the linear kernel as a classifier. Pre-processing is performed by Binarization,Removing the horizontal bar using morphological operations, and then feature extraction followed by classification. MATLAB is used for implementation. The input dataset of the Hindi wordsisgiven as the input. The charactersare extracted from the word and then resized to 70x50 pixels. This resized binary image is divided into 7 horizontal parts.K-means clustering is performedon each part, where the 5 centroids are provided. A

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1827-1830 www.ijaem.net ISSN: 2395-5252

vector with 35 values is created from the image after applying k-means clustering to it. Feature vectorsare created for each character. The Euclidean distance method and the Support Vector Machine (SVM) are used to perform classification, and the outcomes are compared. Results obtained with Support Vector Machine (95.86%) outperform those obtained with Euclidean distance.

In paper[3], the authors used Rectangle Histogram Oriented Gradient[R-HOG] for Feature Extraction and FFANN (feed-forward Artificial Neural Network) and SVM (Support Vector Machine) with RBF(Radial Basis Function) are used for classification. The important objectives for pre-processing used here arenoise reduction, edge detection, connecting tiny broken characters, normalization, region filling, and segmentation. The segmentation is performed using the bounding box. The characters separated are further cropped before being sent for normalization. The experiment is performed using MATLAB 8.0.FFANN(97.15%) performed better than SVM(95.64%) as a classifier.

In the paper [4], the authors performed a comparative study using KNN and SVM classifiers for classifying the handwritten Marathi characters. In pre-processing, noise is removed by using threshold and morphological operations. The segmented characters and the skewed scanned pages are corrected with the help of Hough Transformation technique. The bounding box technique is used for character segmentation from scanned images. The variation in sizes of eachcharacter are normalized to 40 × 40 pixels size. The feature extraction of each character is performed by using the connected pixel-based features like the**eccentricity, area,orientation**, Euler number, and perimeter. The k-nearest neighbour (KNN) and theSVM algorithm along with itslinear and RBF kernelare used for the preparation of the result. The accuracies of the proposed methods by both classifiers are recorded. The Experiment is performed using Matlab 8.0. The overall accuracy obtainedusing KNN is 91.52% and SVM is 95.35%.

In the paper [5], the authors have performed the classification using KNN and SVM for Marathi handwritten characters.The dataset used consists of each Marathi characteralong with different writing styles.The Kaggle dataset is used as the input dataset. The pre-processing is performed by detecting edges with the help ofthe canny method,removing the unwanted regions along with thefillingregions with morphological operations, cropping characters using the bounding box technique followed by normalization of the image. The feature vectors of the pre-processed image are extracted with the help of the HOG method by measuring the direction and gradient of the pixel by using the Sobel filter. The overall accuracy obtained byusing the SVM Algorithm is 95%, whereas the accuracy achieved by using theKNNAlgorithm is 90% on the dataset used for testing.

III. ALGORITHMIC SURVEY

Support Vector Machine (SVM):

[5] authors [2021], gave themathematical model of the proposedproblem in the following way:

They first defined w and the$b$ as the decision hyperplane parameters. The aim of this algorithm is to classify with the maximum marginal hyperplane possible.

In the SVM algorithm, the prediction of data, the unseen features,dependsupon the distancefrom the parameter of the decision hyperplane so that the decision function,

$f(x) = w^t x + b = 0$

In the SVM method, choosing an appropriate hyperplane is a crucial step.

Once the hyperplane is computed,it is then used for predictingthe class of the characters.

The Hypothesis function h(xi),used for the classificationof the group is given as

$$h(xi) = \begin{cases} +1, & \text{if } w.x + b > 0 \\ -1, & \text{if } w.x + b < 0 \end{cases}$$

Characters above the hyperplane have features that fall under class +1, whereascharacters below the hyperplane have features that fall under class .

IV. COMPARATIVE STUDY

4.1 Comparison Table between existingsystems:

Paper	Method1	Accuracy	Method2	Accuracy
[1]	SVM (RBF Kernel)	99.3%	-	-
[2]	SVM	95.86%	Euclidean Distance	81.7%

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1827-1830 www.ijaem.net ISSN: 2395-5252

	(Linear Kernel)			
[3]	SVM (RBF & Linear Kernel)	95.64%	FFANN	97.15%
[4]	SVM (RBF Kernel)	88.53%	KNN	80.25%
[5]	SVM (Linear Kernel)	95%	KNN	90%

Based on training and testing dataset accuracies obtained by the authors are used for creating the comparison table.

V.PROPOSED MODEL:

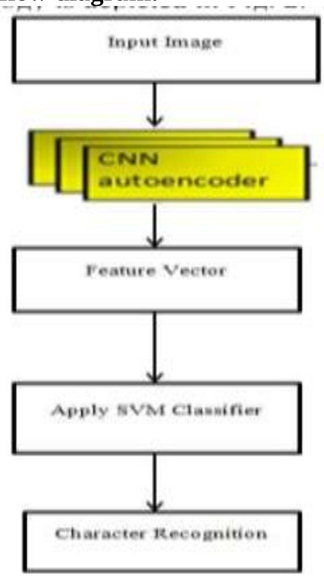
5.1 Introduction:

We have studied different classificationmethods such as:

1. K Nearest Neighbour(KNN)
2. Feedforward Artificial Neural Network(FFANN)
3. Support Vector Machine(SVM)

After studying these three methods, andfinding common findings from thesurveys, we have tried to propose a systemthat will combine the use of SVM usingRBF and Linear Kernel, and CNN Autoencoder as the feature extraction technique. Thesystem will take the Kaggle dataset of handwritten Marathi characters for experiment purposes.The flow diagram of the proposed idea andits related steps to use the proposed model are explained below.

5.2 Proposed flow diagram:



CONCLUSION

In this paper, we've studied how Handwritten characters are identified. We've studied different algorithms used for the classification of the given handwritten characters. As well aswe have studied all the steps from pre-processing to classification. On the basis of this study, we've proposeda system consisting of R-HOG as the feature extraction technique. As we have seen in theabove study, SVM is a betterClassifier over other algorithms for classification purposes. This paper thoroughly studiesdifferent Classifiers used for handwritten Marathi character recognition.

REFERENCES

[1]. S. Joseph and J. George, "Handwritten Character Recognition of MODI Script using Convolutional Neural Network Based Feature Extraction Method and Support Vector Machine Classifier," 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), 2020, pp. 32-36, DOI: 10.1109/ICSIP49896.2020.9339435.

[2]. A. Gaur and S. Yadav, "Handwritten Hindi character recognition using k-means clustering and SVM," 2015 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services, 2015, pp. 65-70, DOI: 10.1109/ETTLIS.2015.7048173.

[3]. Parshuram M. Kamble, Ravindra S. Hegadi, Handwritten Marathi Character Recognition Using R-HOG Feature, Procedia Computer Science, Volume 45, 2015, Pages 266-274, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.03.137.

[4]. Kamble, P.M., Hegadi, R.S. (2017). Comparative Study of Handwritten Marathi Characters Recognition Based on KNN and SVM Classifier. In: Santosh, K., Hangarge, M., Bevilacqua, V., Negi, A. (eds) Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2016. Communications in Computer and Information Science, vol 709. Springer, Singapore. https://doi.org/10.1007/978-981-10-4859-3_9.

[5]. Chikmurge, D., Shriram, R. (2021). Marathi Handwritten Character Recognition Using SVM and KNN Classifier. In: Abraham, A., Shandilya, S., Garcia-Hernandez, L., Varela, M. (eds) Hybrid Intelligent Systems. HIS 2019. Advances in Intelligent Systems and Computing, vol 1179. Springer, Cham. https://doi.org/10.1007/978-3-030-49336-3_32.

[6]. P. K. Singh, S. Das, R. Sarkar, and M. Nasipuri, "Recognition of offline handwritten Devanagari numerals using regional weighted run length features," 2016 International Conference on Computer, Electrical & Communication Engineering (ICCECE), 2016, pp. 1-6, doi: 10.1109/ICCECE.2016.8009567.