xIlkogretim Online - Elementary Education Online,2021; Vol 20 (Issue 5): pp.3733-3744
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.05.410

A Machine Intelligence Based Model for the Classification of Odia Printed and Handwritten Images

Anupama Sahu, Utkal University, Bhubaneswar, Department of Computer Science Engineering and Applications Indira Gandhi Institute of Technology, Sarang, India, anupamasahu.21@gmail.com
Sarojananda Mishra, Department of Computer Science Engineering and Applications Indira Gandhi Institute of Technology, Sarang, India, sarose.mishra@gmail.com
Kalyan Kumar Jena, Department of Computer Science and Engineering, Parala Maharaja Engineering College, Berhampur, India, kalyankumarjena@gmail.com

Abstract: Language plays an important role for the communication among all of us. It is very much essential to detect the printed and handwritten language from several images to extract crucial information from it. In this paper, a machine intelligence (MI) based model is proposed for the classification of Odia printed and handwritten language (OPHL) from the analysis of several Odia language images. The proposed approach is mainly focused on the machine learning (ML) based hybridization mechanism. This mechanism focuses on the combination of the ML based methods such as Logistic Regression (LR) and Neural Network (NN) for the classification of printed and handwritten Odia images. The proposed method is compared with the ML based methods such as Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbour (KNN), AdaBoost (ADB) and Random Forest (RF) methods in terms of Classification Accuracy (CA) performance metric. The proposed method is able to classify the Odia printed and handwritten images in a better way as compared to other methods. The simulation of this work is carried out using Orange 3.26.0.

Keywords: MI, ML, OPHL, SVM, DT, KNN, ADB, RF, LR, NN, CA

I. INTRODUCTION

Language is considered as an important factor for the communication among the human society. Currently, Language recognition and classification is considered as an important research perspectives. The language of entire human society mainly varies state wise of a country and country wise itself. The language can be classified as Indian, Chinese, Russian, Arabic, Spanish, etc. internationally or country wise. By considering the state wise scenario of the country India, the language can be classified as Gujarati, Odia, Tamil, Punjabi, Malayalam, Manipuri, Marathi, Urdu, Hindi, Bengali, etc. But, almost all the countries mainly focus on English language for the communication. In a particular state, the language also varies locally. The language can also be classified as printed and handwritten.

Language recognition and classification is very much essential for its wide variety of applications in the current scenario. Different techniques can be used to classify the languages which can help the blind people for their effective communication.So, it is very much essential to recognize and classify the languages properly as it is very much essential for the effective communication in the real world scenario.

Several research works have been carried out for the classification of languages [1-22]. MI plays an important role the classification of languages. Different ML [23] based methods such as SVM, DT, KNN, ADB, RF etc. can be used for such classification. In this work, we have proposed a model for the classification of Odia printed and handwritten languages from the analysis of several OPHL images. The proposed approach is focused on the combination of LR and NN MI based methods to carry out such classification.

The main contribution of this paper is described as follows.

- A MI based model is proposed for the classification of OPHL images into printed and handwritten category from the analysis of several OPHL images.
- This proposed approach is focused on the combination of LR and NN methods to carry out such classification.

- The proposed method provides better classification results as compared to the existing methods such as SVM, DT, KNN, ADB and RF in terms of CA.
- The proposed work is carried out using Orange 3.26.0.

The rest of the paper is organized as follows. Section II, III, IV and V describes the related works, proposed model, results and discussion,and conclusion of this work.

II. RELATED WORKS

Different works have been carried out by several researchers for Odia character processing and analysis [1-22]. Some of the works are described as follows. Singh et al. [1] emphasized on Odia character recognition mechanism by focusing on feature extraction and classification techniques. Dash et al. [2] focused on the handwritten Odia numeral recognition using a hybrid feature and discriminant classifier. Sethy et al. [3] emphasized on the offline handwritten Odia character recognition using DWT and PCA technique. Nayak et al. [5] focused on the backpropagation network with binary features for the recognition of Odia character. Padhi et al. [9] emphasized on Odia handwritten character recognition by using zone centroid distance and standard deviation based feature matrix.

Das et al. [12] focused on the Odia handwritten numeral recognition by emphasizing on LSTM mechanism. Satpathy et al. [18] emphasized on the printed Odia numeral recognition by the help of stacked autoencoder mechanism. Sahu et al. [19] focused on the Odia script analysis of several recognition techniques by emphasizing on classification mechanism and their accuracy measure. Mohapatra et al. [20] emphasized on two stage classification mechanism to recognize handwritten atomic Odia character. Pujari et al. [21] focused on Slantlet Transform and differential evolution based functional link artificial NN classifier for Odia handwritten vowel recognition. Mishra et al. [22] emphasized on handwritten Odia numerals recognition by using a contour descriptors based generalized scheme.

III. PROPOSED MODEL

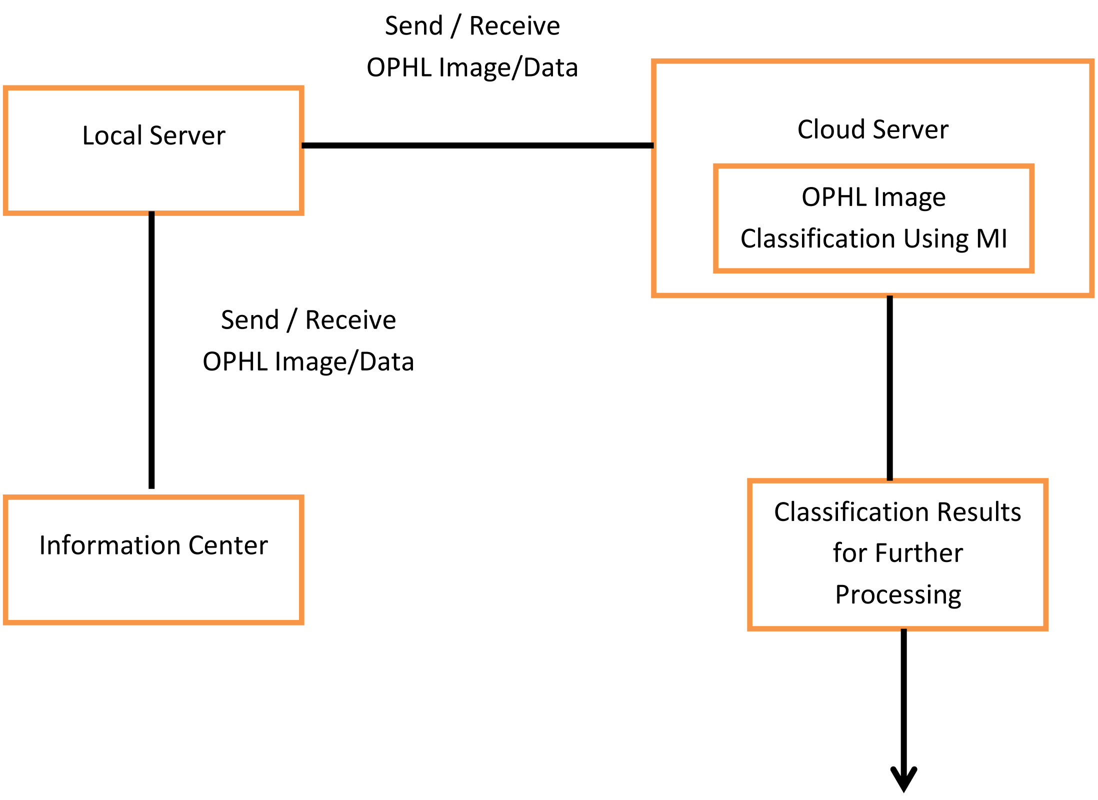

Fig.1 Proposed Model

The proposed model is described in Fig. 1. In this work, it is assumed that the OPHL images will be stored in an information center (IC). From the IC the OPHL images will be sent to the local server. As the local server is not able to process the OPHL images due to the lack of computing mechanism, so from the local server, the OPHL images will be sent to the cloud server for their processing. In the cloud server, the OPHL images will be processed using the proposed method to provide the classification results for further processing. In this model, the main focus is given on the classification of OPHL images using proposed method and other existing ML [23] based methods such as SVM, DT, KNN, ADB and RF using MI mechanism. The proposed methodology is described in Fig. 2.

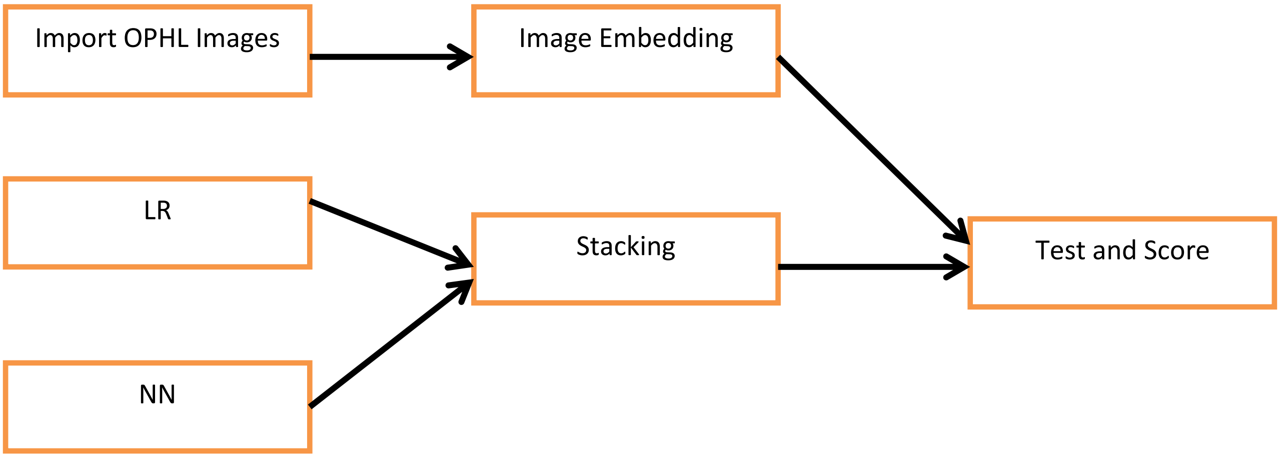

Fig. 2 Proposed Methodology

The proposed methodology is mainly focused on the hybridization of the ML [23] based methods such as LR and NN. The set up mechanism for the LR and NN methods are described in Table 1 and Table 2 respectively. For LR, we have considered Ridge (L2) and C=1 as mentioned in Table 1. For NN, we have considered ReLu activation function and Adam solver as mentioned in Table 2. In our work, initially the OPHL images are provided to the Orange 3.26.0 [24].Then, the image embedding mechanism is carried out. The LR and NN methods are hybridized using Stacking mechanism. The CA is computed using Test and Score by focusing on image embedding and Stacking mechanism. The proposed method is compared with SVM, DT, KNN, ADB and RF methods.

Table 1 Set up for LR Method for Proposed Methodology

Parameter	Assigned Value
Regularization Type	Ridge (L2)
Strength	C = 1

Table 2 Set up for NN Method for Proposed Methodology

Parameter	Assigned Value
Neurons in Hidden Layers	100
Activation	ReLu
Solver	Adam
Regularization	a = 0.0001
Maximum Number of Iterations	200
Replicable Training	Yes

The set up mechanism for SVM, DT, KNN, ADB and RF methods are described in Table 3 to Table 7 respectively. For SVM, we have considered linear kernel and SVM type processing as mentioned in Table 3. For DT, we have considered the inclusion of binary tree and have focused on the 2 number of instances in leaves as mentioned in Table 4. For KNN, we have considered Manhattan metric and distance weight as mentioned in Table 5. For ADB, we have consideredSAMME.R classification algorithm as mentioned in Table 6. For RF, we have considered 20 number of trees for processing as mentioned in Table 7.

Table 3 Set up for SVM Method

Parameter	Assigned Value
SVM Type	SVM
Cost (C)	1.00
Regression Loss Epsilon	0.10
Kernel	Linear
Numerical Tolerance	0.0010
Iteration Limit	100

Table 4 Set up for DT Method

Parameter	Assigned Value
Minimum Number of Instances in Leaves	2
Do not Split Subsets Smaller Than	5
Limit the Maximal Tree Depth to	100
Stop When Majority Reaches [%]	95
Include Binary Tree	Yes

Table 5 Set up for KNN Method

Parameter	Assigned Value
Number of Neighbors	5
Metric	Manhattan
Weight	Distance

Table 6 Set up for ADB Method

Parameter/Algorithm	Assigned Value/Algorithm/Function
Base Estimator	Tree
Number of Estimators	50
Learning Rate	1.00000
Classification Algorithm	SAMME.R
Regression Loss Function	Linear

Table 7Set up for RF Method

Parameter	Assigned Value
Number of Trees	20
Number of Attributes Considered at Each Split	5
Limit Depth of Individual Trees	3
Do not Split Subsets Smaller Than	10

IV. RESULTS AND DISCUSSION

In this work, Orange 3.26.0 [24] is used to carry out the classification of OHPL images. Here, we have processed several Odia printed language images [25-34] and Odia handwritten language images of same Odia printed language images which are mentioned in Fig. 3 and Fig. 4. The Orange classification framework is mentioned in Fig. 5. The results are mentioned in Table 8 to Table 10 and Fig. 6 to Fig. 11by focusing on the CA both in 0-1 unit and in percentage format.

Fig. 3 Odia Printed Language Images

Fig. 4 Odia Handwritten Language Images

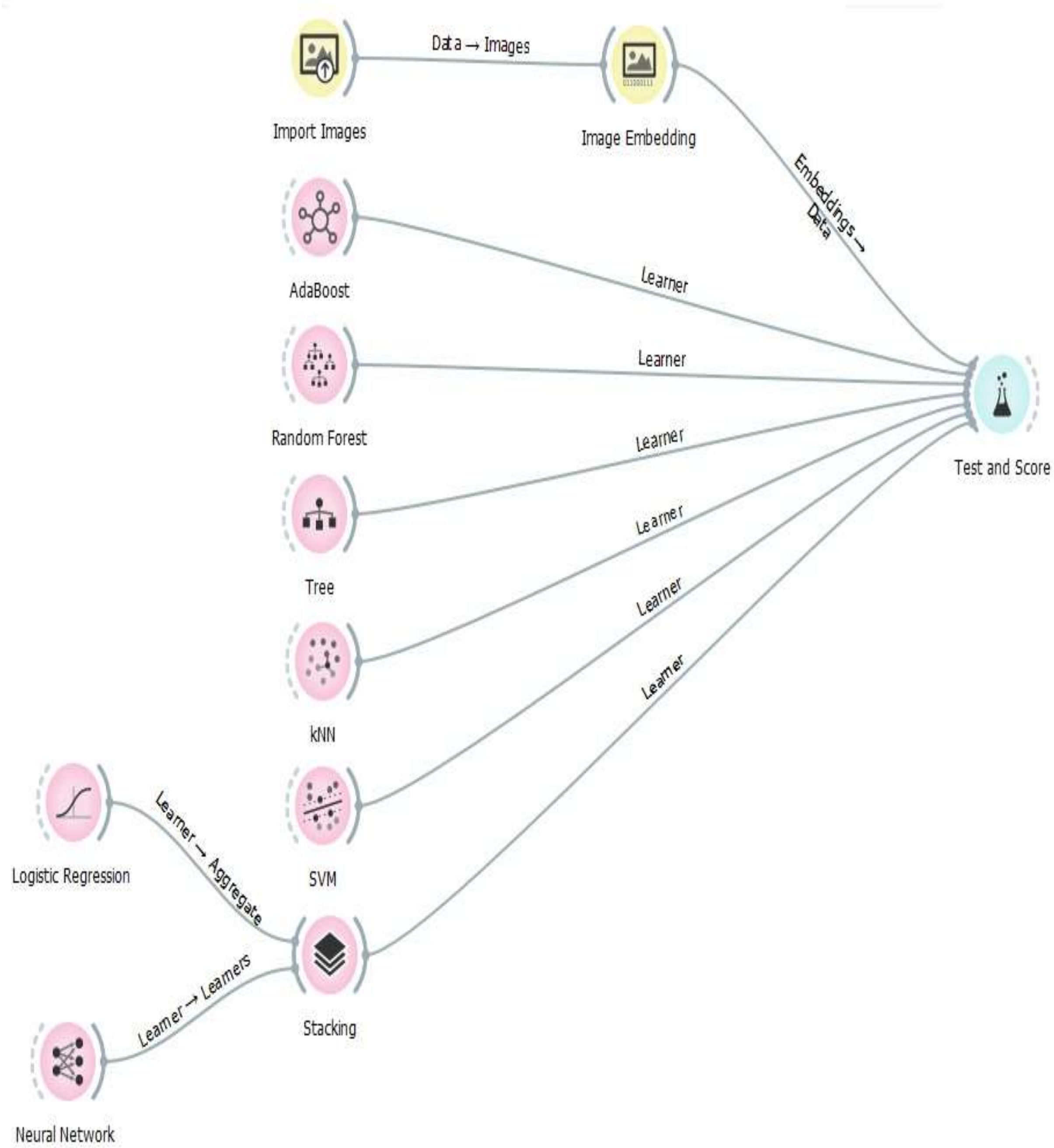

Fig. 5 Classification Framework

Case I: (NOFDS=2)

Table 8 Performance evaluation of Proposed and other methods when NOFDS=2

Method	CA (in 0-1 unit)	CA (in %)
SVM	0.900	90
DT	0.800	80
KNN	0.700	70
ADB	0.700	70
RF	0.500	50

Stack / Proposed Method	0.950	95

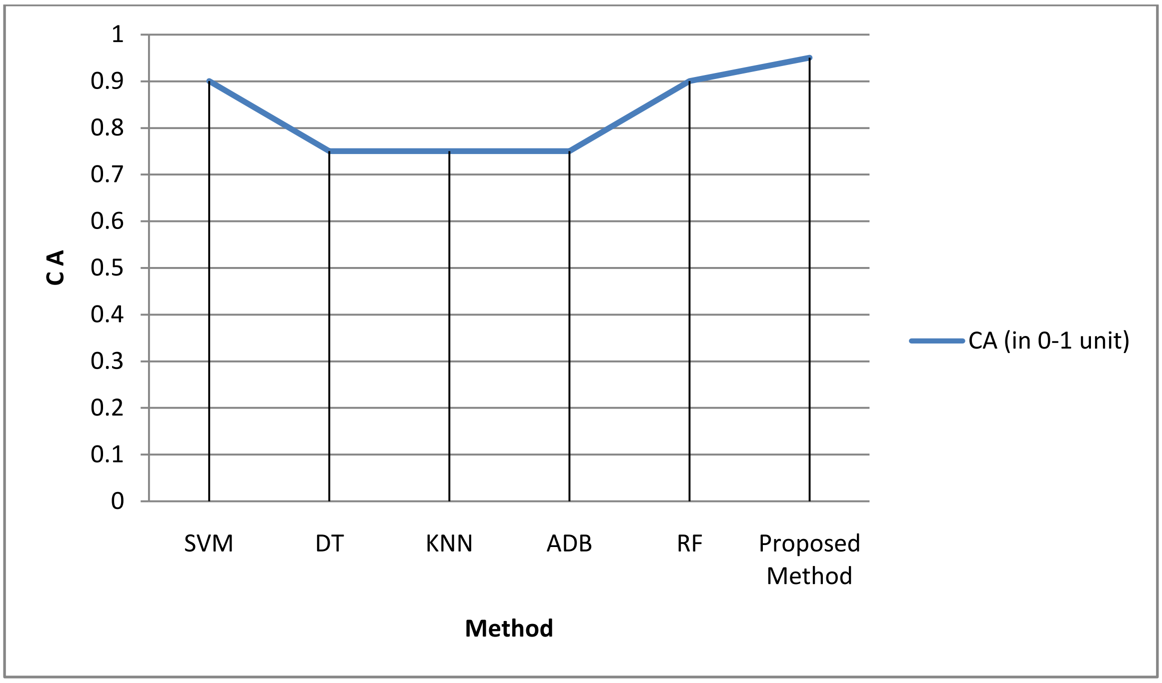

Fig. 6 CA of different methods in 0-1 unit when NOFDS=2

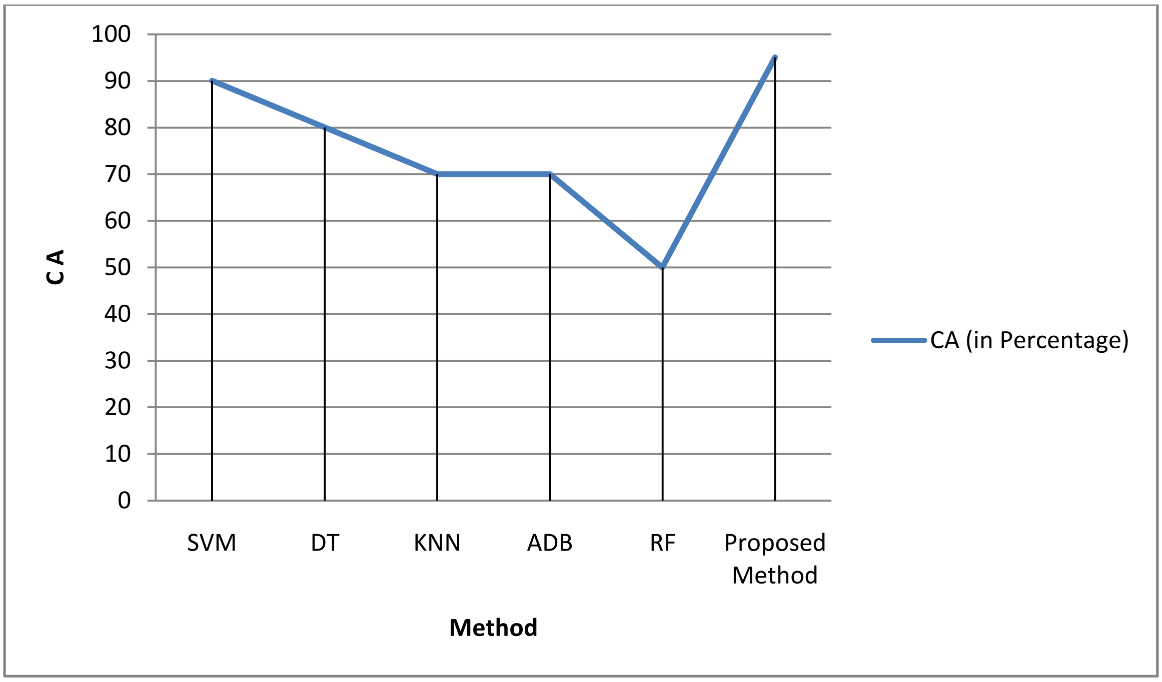

Fig. 7 CA of different methods in percentage when NOFDS=2

Case II: (NOFDS=3)

Table 9 Performance evaluation of Proposed and other methods when NOFDS=3

Method	CA (in 0-1 unit)	CA (in %)
SVM	0.900	90
DT	0.750	75
KNN	0.750	75
ADB	0.750	75
RF	0.900	90
Stack / Proposed Method	0.950	95

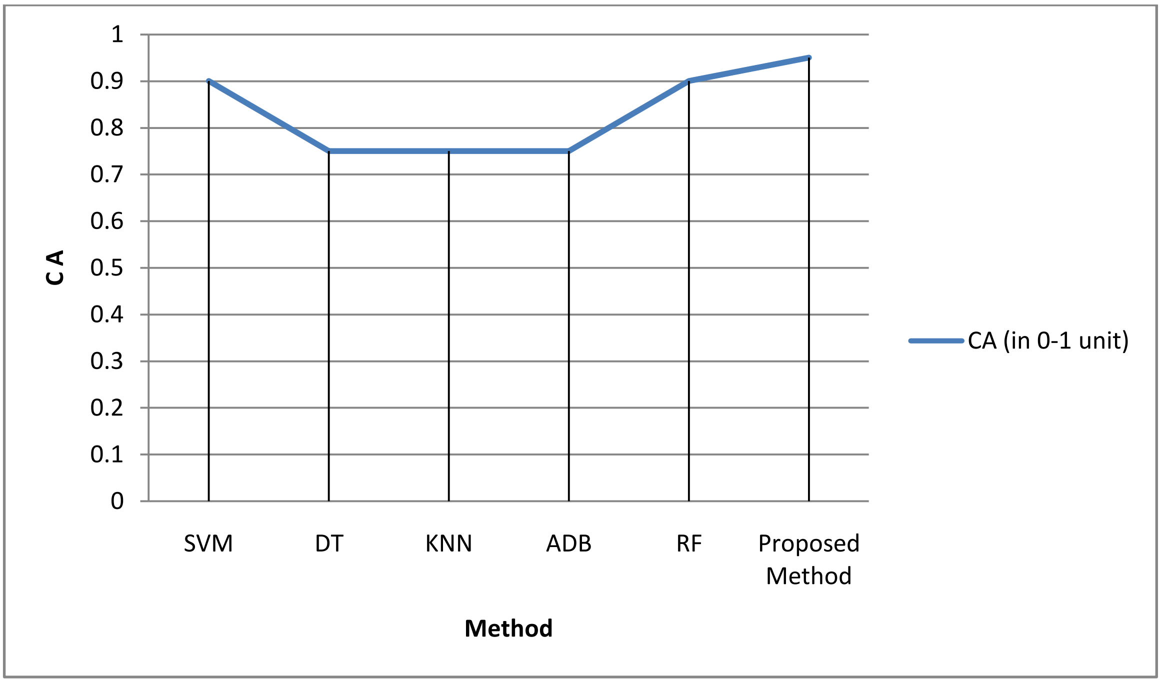

Fig. 8 CA of different methods in 0-1 unit when NOFDS=3

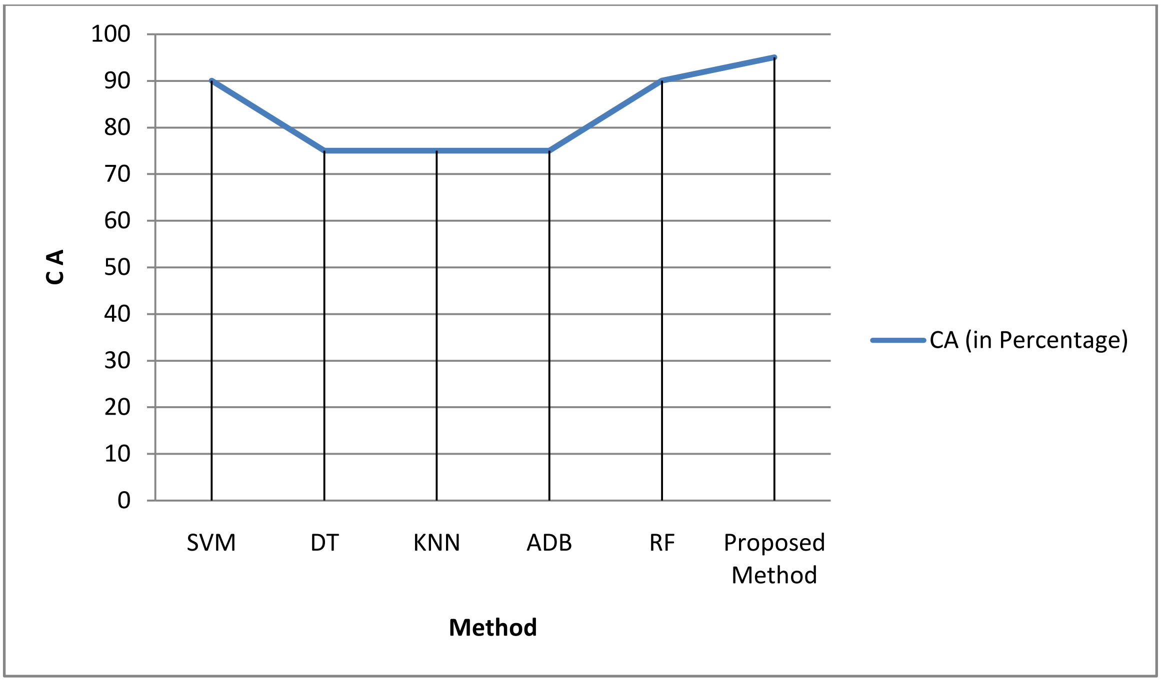

Fig. 9 CA of different methods in percentage when NOFDS=3

Case III: (NOFDS=5)

Table 10 Performance evaluation of Proposed and other methods when NOFDS=5

Method	CA (in 0-1 unit)	CA (in %)
SVM	0.900	90
DT	0.550	55
KNN	0.900	90
ADB	0.650	65
RF	0.900	90
Stack / Proposed Method	0.950	95

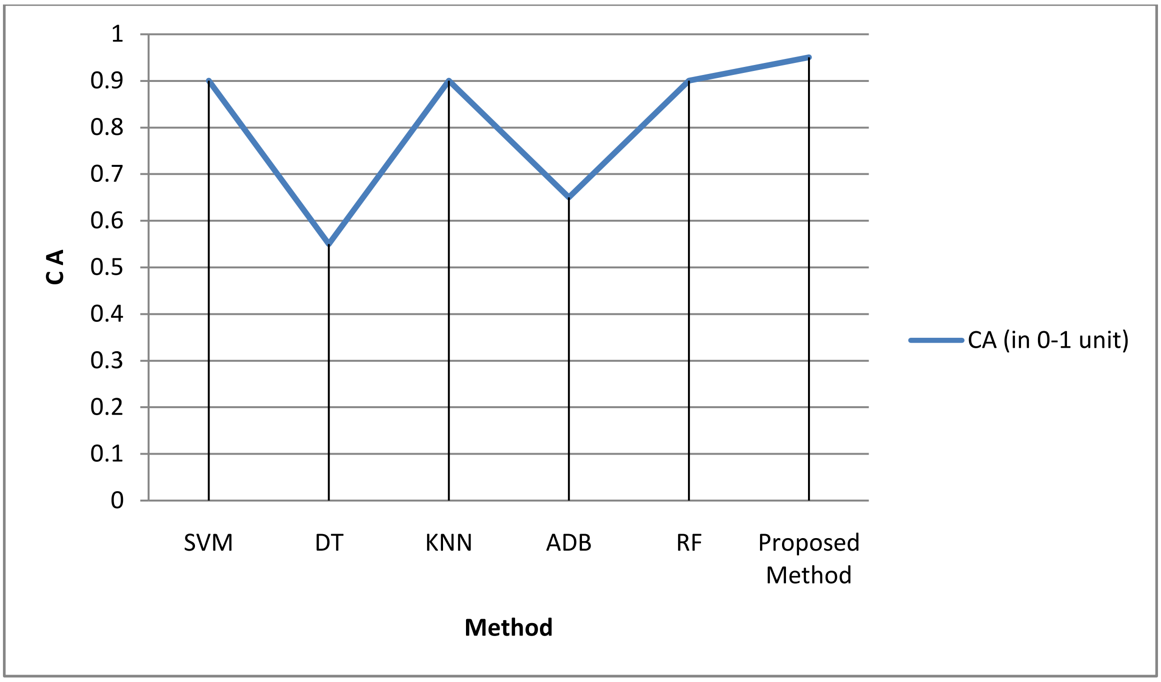

Fig. 10 CA of different methods in 0-1 unit when NOFDS=5

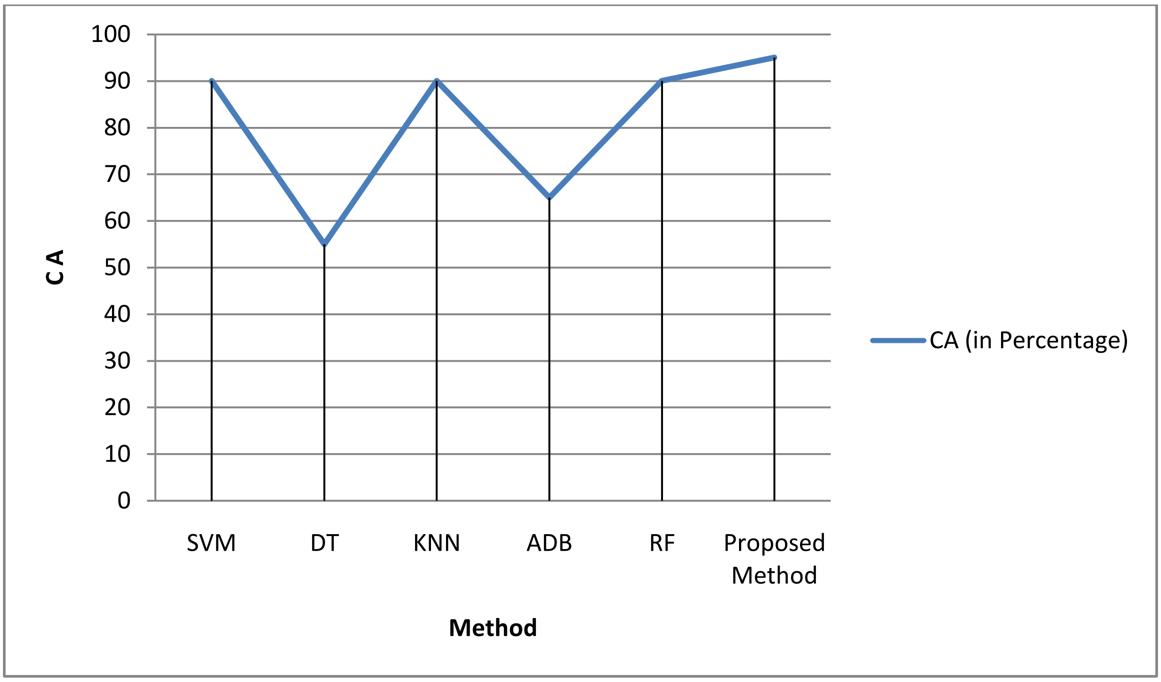

Fig. 11 CA of different methods in percentage when NOFDS=5

From the results, it is observed that, the proposed method is able to provide better classification results with 95% (0.950 in 0-1 unit) of CA in all the mentioned cases as compared existing methods such as SVM, DT, KNN, ADB, RF methods. When the NOFDS is considered as 2, then the SVM, DT, KNN, ADB, RF methods provide 90, 80, 70, 70, 50 CA (in percentage) or 0.900, 0.800, 0.700, 0.700, 0.500 CA (in 0-1 unit) respectively. When the NOFDS is considered as 3, then the SVM, DT, KNN, ADB, RF methods provide 90, 75, 75, 75, 90 CA (in percentage) or 0.900, 0.750, 0.750, 0.750, 0.900 CA (in 0-1 unit) respectively. When the NOFDS is considered as 5, then the SVM, DT, KNN, ADB, RF methods provide 90, 55, 90, 65, 90 CA (in percentage) or 0.900, 0.550, 0.900, 0.650, 0.900 CA (in 0-1 unit) respectively.

So, the existing method SVM is able to provide 90% CA which is maximum and the existing method RF is able to provide 55% CA which is minimum as compared to the other existing methods in the scenario of NOFDS=2. Similarly, the methods SVM and RF is able to provide 90% CA which is maximum and the methods DT, KNN and ADB is able to provide 75% CA which is minimum by considering the existing methods in the scenario of NOFDS=3. The methods SVM, KNN and RF is able to provide 90% CA which is maximum and the method DT is able to provide 55% CA which is minimum by considering the existing methods in the scenario of NOFDS=5.The existing methods are able to provide maximum 90% CA by considering the mentioned scenario. However, the proposed method is able to provide 95% in all the mentioned scenarios which is maximum as compared to other existing mentioned methods.

V. CONCLUSION

This paper proposed a MI based model for the classification of OPHL from the analysis of several Odia language images. The proposed approach is focused on the hybridization of LR and NN methods for the classification of printed and handwritten Odia images. The proposed method is compared with SVM, DT, KNN, ADB, RF methods in terms of CA. From the results, it is concluded that the proposed method is able to classify the Odia printed and handwritten images with 95% CA in all the cases such as NOFDS=2, NOFDS=3 and NOFDS=5. So, the proposed method performs better in all the cases in terms of CA as compared to other methods. This work can be extended to develop improved methods to classify the OPHL images and other images in a better way.

REFERENCES

1. Singh, S., Sarangi, P. K., Singla, C., &Sahoo, A. K. (2020). Odia character recognition system: A study on feature extraction and classification techniques. *Materials Today: Proceedings*.
2. Dash, K. S., Puhan, N. B., & Panda, G. (2014, April). A hybrid feature and discriminant classifier for high accuracy handwrittenOdia numeral recognition. In *2014 IEEE REGION 10 SYMPOSIUM* (pp. 531-535). IEEE.
3. Sethy, A., Patra, P. K., &Nayak, D. R. (2018). Off-line handwritten Odia character recognition using DWT and PCA. In *Progress in Advanced Computing and Intelligent Engineering* (pp. 187-195). Springer, Singapore.
4. Pattanayak, S. S., Pradhan, S. K., &Mallik, R. C. (2020). Printed Odia Symbols for Character Recognition: A Database Study. In *Advanced Computing and Intelligent Engineering* (pp. 297-307). Springer, Singapore.
5. Nayak, M., &Nayak, A. K. (2017). Odia character recognition using backpropagation network with binary features. *International Journal of Computational Vision and Robotics*, *7*(5), 588-604.
6. Sethy, A., &Patra, P. K. (2020). R-HOG Feature-Based Off-Line Odia Handwritten Character Recognition. In *Examining Fractal Image Processing and Analysis* (pp. 196-210). IGI Global.
7. Dash, K. S., Puhan, N. B., & Panda, G. (2017). Odia character recognition: a directional review. *Artificial Intelligence Review*, *48*(4), 473-497.
8. Puri, S., & Singh, S. P. (2020). A fuzzy matching based image classification system for printed and handwritten text documents. *Journal of Information Technology Research (JITR)*, *13*(2), 155-194.
9. Padhi, D., &Senapati, D. (2013). Zone centroid distance and standard deviation based feature matrix for odia handwritten character recognition. In *Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)* (pp. 649-658). Springer, Berlin, Heidelberg.
10. Mishra, T. K., Majhi, B., & Panda, S. (2013, August). A comparative analysis of image transformations for handwritten Odia numeral recognition. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 790-793). IEEE.
11. Sethy, A., &Patra, P. K. (2016, April). Off-line Odia handwritten numeral recognition using neural network: a comparative analysis. In *2016 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 1099-1103). IEEE.
12. Das, A., Patra, G. R., &Mohanty, M. N. (2020, July). LSTM based Odia Handwritten Numeral Recognition. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0538-0541). IEEE.

13. Das, A., Patra, G. R., &Mohanty, M. N. (2021). A Comparison Study of Recurrent Neural Networks in Recognition of Handwritten Odia Numerals. In *Advances in Electronics, Communication and Computing* (pp. 251-260). Springer, Singapore.
14. Sahoo, R. C., &Pradhan, S. K. (2021). Pattern Storage and Recalling Analysis of Hopfield Network for Handwritten Odia Characters Using HOG. In *Advances in Machine Learning and Computational Intelligence* (pp. 467-476). Springer, Singapore.
15. Sethy, A., Patra, P. K., &Nayak, S. R. (2020). Offline Handwritten Numeral Recognition Using Convolution Neural Network. *Machine Vision Inspection Systems: Image Processing, Concepts, Methodologies and Applications*, *1*, 197-212.
16. Behera, S., & Das, N. (2020). Handwritten Odia Numerals Recognition: A Supervised Learning Perspective. In *Advances in Intelligent Computing and Communication* (pp. 86-93). Springer, Singapore.
17. Nayak, S., Pradhan, S. K., Biswal, P. K., & Jena, O. P. (2019, November). Classification Performance Analysis of Printed Odia Alphabets using Neural Network based Approach. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 176-180). IEEE.
18. Satpathy, S., Nayak, A. K., Nayak, M., & Patnaik, S. (2019). Printed Odia Numeral Recognition Using Stacked Autoencoder. *Procedia Computer Science*, *154*, 790-797.
19. Sahu, R. K., Minz, A., Mishra, S., & Panda, C. (2017). Analysis of Different Recognition Technique Classification and their Accuracy Measure over Odia Script-A Review.
20. Mohapatra, R. K., Majhi, B., & Jena, S. K. (2017). Two-stage classification approach for the recognition of handwritten atomic Odia character. *International Journal of Applied Pattern Recognition*, *4*(1), 44-63.
21. Pujari, P., &Majhi, B. (2017). Odia Handwritten Vowel Recognition System using Slantlet Transform and Differential Evolution based Functional Link Artificial Neural Network Classifier. *International Journal of Advanced Research in Computer Science*, *8*(5).
22. Mishra, T. K., Majhi, B., & Dash, R. (2017). A Contour Descriptors-Based Generalized Scheme for Handwritten Odia Numerals Recognition. *JIPS*, *13*(1), 174-183.
23. Sen, Pratap Chandra, MahimarnabHajra, and MitadruGhosh, "Supervised classification algorithms in machine learning: A survey and review," In *Emerging Technology in Modelling and Graphics*, pp. 99-111. Springer, 2020.
24. https://orange.biolab.si/download/#windows
25. http://incredibleodisha.blogspot.com/2009/10/odia-our-mother-tongue.html
26. https://en.wikipedia.org/wiki/Odia_language
27. https://commons.wikimedia.org/wiki/File:Oriya_Lipi.svg
28. https://www.quora.com/Are-you-proud-to-be-an-Odia
29. https://odishanewsinsight.com/odisha/odia-becomes-official-language-in-odisha-from-today/
30. https://incredibleorissa.com/oriya-news-papers/
31. https://androdinewsmagazinesapps.wordpress.com/2015/11/20/daily-odia-news-daily-odia-news-oriya-news-odisha-khabar-dailyhunt/
32. https://en.wikipedia.org/wiki/Sambalpuri_language
33. https://radiosindia.com/airodia.html
34. https://steprimo.com/android/us/app/com.atuldalnaik.Odia_Speech_Book/Speech-and-Essay-in-Odia/