



## HYBRID MACHINE LEARNING IN CLASSIFICATION METHODS FOR HCR IN GUJARATI LANGUAGE

PRIYANK D. DOSHI

ATMIYA UNIVERSITY,RAJKOT-360005, GUJARAT, INDIA

DR.PRATIK VANJARA

DEPARTMENT OF COMPUTER SCIENCE SHREE MANIBHAI VIRANI & SMT. NAVALBEN  
VIRANI SCIENCE COLLEGE, RAJKOT-360005, GUJARAT, INDIA

### ABSTRACT

The problem of recognizing Gujarati Handwritten character with vowels opening new future scope where one can use smart phone, website or any handy scanner to convert hand written Gujarati Language into text. It will be very effective to give education in mother language at primary level. Public, Private and Government sectors will be benefited when they get any hand written Gujarati Script and they can directly convert it into softcopy or into text form. There are many methods used to solve this problem.Using CNN we can improve new algorithm depending on training data set, mathematical model and other intricacy. Convolutional Neural Network or machine learning is very useful for this. Still there are more chances for improvement and rising accuracy using Machine learning in combination of Deep Learning as a hybrid model.

**KEYWORDS:***MACHINE LEARNING, DEEP LEARNING, CONVOLUTIONAL NEURAL NETWORK, SUPPORT VECTOR MACHINE, HAND WRITTEN CHARACTER RECOGNITION (HCR), ARTIFICIAL NEURAL NETWORK (ANN).*

### INTRODUCTION

As we know retyping text at the point of any transaction can cause errors. HCR for Gujarati is very effective as it has many advantages in schools, government offices, and public and private sector offices.

In the solution model we have to follow certain steps like stages like Image Acquisition, Pre-processing, Segmentation, Feature extraction, classification and post processing. Using smart phones, websites and hand scanner tools we can put the solution online. That will give ease to every person who wants to communicate in Gujarati at places like government record office, courier office, post offices and involved in document conversion and preservations.

The future scope will be applicability of algorithm by using it public domain using internet. Our Urban and Rural areas of Gujarat are well developed with internet availability and Internet of things. So starting from schools where we need to implement NEP 2020, we require to make it useful for each and every person. Once the problem of hand written Gujarati characters with vowels recognition solved, future scope will be detection and conversion of Gujarati phonetic language to text.

Machine Learning and Deep Learning in hybrid way as a classifier can be novel approach to solve the problem of HCR.

**HYBRID MODEL**

As shown in below figure it comprised stages like Image Acquirement, Pre-processing, Segmentation, Feature extraction, classification and post processing. Deep Learning and machine learning both can be applied separately but Hybrid Model may ensemble different feature extraction techniques and classification techniques of ML or DL. As shown in Fig. 1. we can develop model that ensemble both ML and DL.

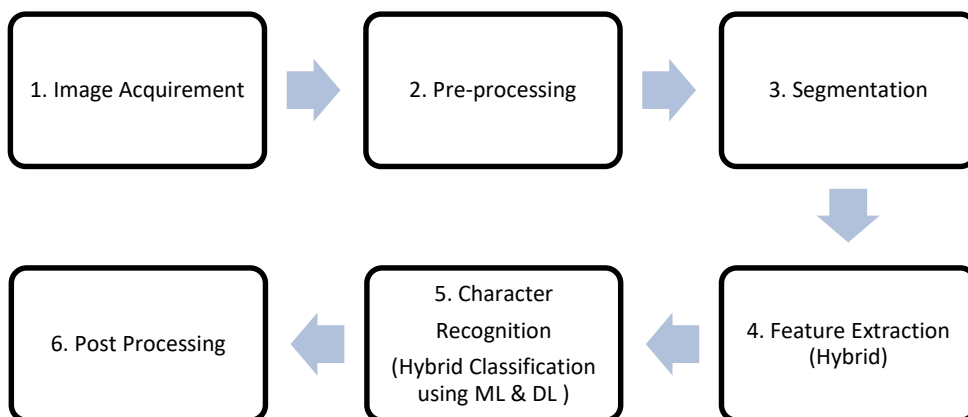


Figure 1 Hybrid Model using ML & DL

**1 Comparison of DL & ML**

Factors	Deep Learning	Machine Learning
Size of Image Set	High	Comparatively Low
Accuracy	High	Comparatively Low
Computational Cost		
(Training Time)	Longer	Comparatively Short
(Hardware Dependency)	GPU required	Can Train on CPU
Hyper Parameter Tuning	Comparatively More	Limited

**ML ALGORITHM**

To rise efficiency and accuracy in machine learning fine tuning of parameters is required. Data set creation, Training of data set, Evaluation Methods and model types are such factors using which we can improve ML algorithm.

**DATA SET CREATION**

As shown in below given figure, Hand written images are segmented in 28 by 28 pixels images. Since there are twelve vowels and thirty six characters we can have 432 distinct combined characters. At maximum 432 combined characters class we require with sufficient image set in each class depending on ML or DL are used. Deep learning approach is more hungry for dataset.

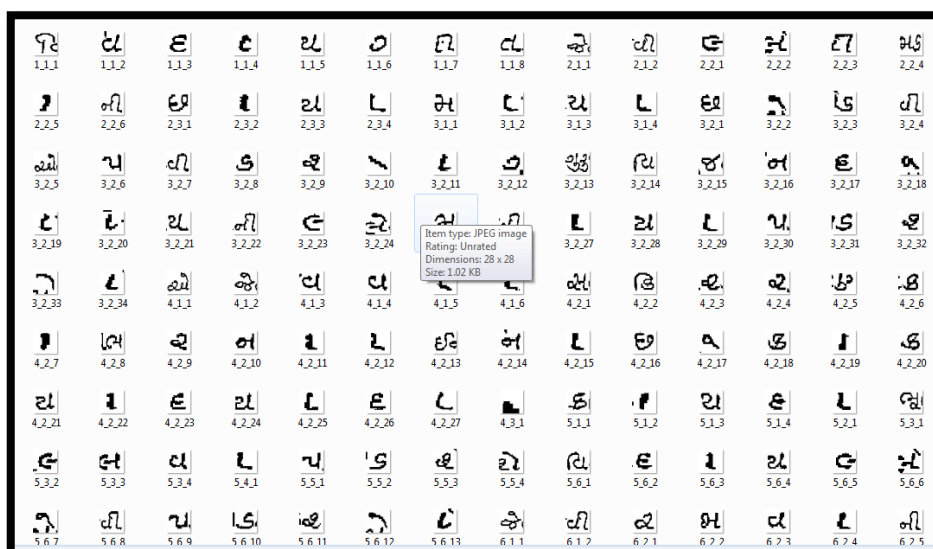


Figure 2: Segmented Characters of Gujarati Language

**FEATURE EXTRACTION**

There is no need for feature extraction when using a deep network as one can just pass the images directly to the network and generally achieve good performance. This completely eliminates the big and grueling point engineering stage of the whole process. It rises network layers to perform feature extraction itself.

Classical ML algorithms frequently bear complex point engineering. Generally, a deep dive exploratory data analysis is first performed on the image set to extract features. A dimensionality reduction might also be done for easier processing. Finally, features must be precisely named to pass over to the ML algorithm.

**REVIEW ON CLASSIFICATION OF METHODS**

For offline recognition of Segmented Gurumukhi characters Anupam Garg, Manish kumar Jindal and Amarpreet Singh presented PCA and modified division point-based features and giving them as input in classification process k-NN and SVM. It has three kernels linear, polynomial and RBF – SVM. Data set size was 8960 samples of offline handwritten Gurumukhi characters. They got 92.3 % accuracy. Gujrati Vowel was not included in this. So it grabs the interest to do research for that[1].

A detailed survey with timeline on support vector machine published by Mayank Arya Chandra and S. S. Bedi. They have included various techniques about support vector machine. In their survey they have explained analysis of SVM as well as neural network[2].

Machine learning based model was presented by Abid Sarwar, Mehboob Ali, Jatinder Manhas and Vinod Sharma. In their work they have solved the problem of diagnosis of diabetes type – II. Data set collected from the group of 400 people. They have used Artificial Neural Networks, K-Nearest Neighbours, Navie Bayes and support vector Machine to make it hybrid. Result of each one ensembled and final result was provided[3].

## RESEARCH MATRIX :2321-7073

*Peer Reviewed & Refereed | International Multidisciplinary journal of applied research*

---

To recognize handwritten Marathi numeral Deepak T. Mane, Rushikesh Tapdiya and Swati V. Shinde developed model that was based on ensemble neural network. Using CNN, they have created 5 base pipelines and then its output goes to meta classifier network to get desired output. Whole work was based on machine learning so they used 63000 samples for training, 7000 samples used for validation and 11500 samples used for testing. They concluded and suggested to use dilated customized CNN to increase accuracy and reduce computational cost[4].

Shailesh Acharya, Ashokkumar Pant and Prashanna kumar Gyawali presented a new dataset for Devanagari Handwritten recognition problem. It consists 92 thousand images of 46 different characters of Devanagari script. We explored the challenges in classification of characters in Devanagari Dataset. Combined characters specially with vowels are not included in this research. They used deep learning models for recognition. The experimental results shows 98.47% accuracy[5].

A paper titled “A Review of Deep Learning in Image Recognition” presented by Myeongsuk Pak and Sanghoon Kim explains that “Deep learning has gained tremendous progress because it uses big data, powerful computer, and algorithms. Convolutional neural network is one of the most prominent approaches. Convolutional neural network has won the ILSVRC in recent years. It has outstanding results in image recognition. The convolutional neural networks achieved high accuracy with significantly reduced error rate.[6]”

According to Dhara S. Joshi and Yogesh R. Risodkar presented in the paper titled “Deep Learning Based Gujarati Handwritten Character Recognition” that “Deep learning is the method used for identity recognition which will depend on the man-made neural network. Deep learning involves a proper focus on right features of the image to recognize themselves with a negligible assistance from the program” [7].

S. Liaqat, K. Dashtipour, K. Arshad, K. Assaleh, and N. Ramzan used hybrid approach of deep learning and machine learning in Posture detection problem [8]. They have applied support vector machine (SVM), logistic regression (KNN), decision tree, Naive Bayes, random forest, Linear discrete analysis and Quadratic discrete analysis) and deep learning classifiers (i.e., 1D-convolutional neural network (D-CNN), 2D-convolutional neural network (2D-CNN), LSTM and bidirectional LSTM) to identify posture detection. Ultimately, they can predict about health of different age of people[8].

A detailed survey covering all techniques relevant to Gujarati character recognition problem related to feature extraction to classification has been performed by Priyank D. Doshi and Pratik A. Vanjara where we can see that machine learning, deep learning is more potential to achieve the goal so it is well understood to use both these technologies[9].

As we have seen above, Hybrid concept is successful in feature extraction and also in classifier modelling. We can apply hybrid concept using deep and machine learning to get more robust algorithm for Gujarati Character with vowels recognition.

### CONCLUSIONS

Fusion or ensemble of multiple approaches either in feature extraction or in classifier will be the future in probable solutions for Gujarati character with vowels recognition problem. As we know both has their pros and cons and tuning their parameters collectively to get benefits from both of them surprisingly good results are observed. In this paper several classifier and

VOLUME-1 / YEAR -9 / ISSUE -6 / JANUARY - 2022  
WWW.RESEARCHMATRIX.ORG

## RESEARCH MATRIX :2321-7073

*Peer Reviewed & Refereed | International Multidisciplinary journal of applied research*

feature extraction techniques are reviewed corresponding to feature extraction and classifier for character recognition. In conclusion hybrid option will be the best with sufficient data set for Gujarati Language with appropriate intricacy of machine learning along with deep learning.

### REFERENCES

- [1] A. Garg, M. K. Jindal, and A. Singh, "Offline handwritten Gurmukhi character recognition: k-NN vs. SVM classifier," *Int. J. Inf. Technol.*, vol. 13, no. 6, pp. 2389–2396, 2021, doi: 10.1007/s41870-019-00398-4.
- [2] M. A. Chandra and S. S. Bedi, "Survey on SVM and their application in image classification," *Int. J. Inf. Technol.*, vol. 13, no. 5, pp. 1867–1877, 2021, doi: 10.1007/s41870-017-0080-1.
- [3] A. Sarwar, M. Ali, J. Manhas, and V. Sharma, "Diagnosis of diabetes type-II using hybrid machine learning based ensemble model," *Int. J. Inf. Technol.*, vol. 12, no. 2, pp. 419–428, 2020, doi: 10.1007/s41870-018-0270-5.
- [4] D. T. Mane, "Handwritten Marathi numeral recognition using stacked ensemble neural network," vol. 13, no. October, pp. 1993–1999, 2021.
- [5] S. Acharya, A. K. Pant, and P. K. Gyawali, "Deep learning based large scale handwritten Devanagari character recognition," *Ski. 2015 - 9th Int. Conf. Software, Knowledge, Inf. Manag. Appl.*, 2016, doi: 10.1109/SKIMA.2015.7400041.
- [6] M. Pak, S. Kim, and A. Alexnet, "A Review of Deep Learning in Image Recognition," pp. 2013–2015, 2014.
- [7] D. S. Joshi and Y. R. Risodkar, "Deep Learning Based Gujarati Handwritten Character Recognition," *2018 Int. Conf. Adv. Commun. Comput. Technol. ICACCT 2018*, pp. 563–566, 2018, doi: 10.1109/ICACCT.2018.8529410.
- [8] S. Liaqat, K. Dashtipour, K. Arshad, K. Assaleh, and N. Ramzan, "A Hybrid Posture Detection Framework: Integrating Machine Learning and Deep Neural Networks," *IEEE Sens. J.*, vol. 21, no. 7, pp. 9515–9522, 2021, doi: 10.1109/JSEN.2021.3055898.
- [9] P. D. Doshi, P. A. Vanjara, "A Comprehensive Survey on Handwritten Gujarati Character and Its Modifier Recognition Methods." In: Joshi A., Mahmud M., Ragel R.G., Thakur N.V. (eds) *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Lecture Notes in Networks and Systems, vol 191. Springer, Singapore. [https://doi.org/10.1007/978-981-16-0739-4\\_79](https://doi.org/10.1007/978-981-16-0739-4_79)