





Enhancing Marathi Handwritten Character Recognition Using Ensemble Learning

Diptee Vishwanath Chikmurge^{1,2*}, Shriram Raghunathan¹

¹ School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh 466114, India

² School of Computer Engineering, MIT Academy of Engineering, Alandi, Maharashtra 412105, India

Corresponding Author Email: chdiptee@gmail.com

<https://doi.org/10.18280/ts.400132>

Received: 9 October 2022

Accepted: 6 February 2023

Keywords:

AlexNet, Lenet-5, VGG-16, ensemble learning, stacking, and voting

ABSTRACT

In the current era, the Optical Character Recognition (OCR) model plays a vital role in converting images of handwritten characters or words into text editable script. During the COVID-19 pandemic, students' performance is assessed based on multiple-choice questions and handwritten answers so, in this situation, the need for handwritten recognition has become acute. Handwritten answers in any regional language need the OCR model to transform the readable machine-encoded text for automatic assessment which will reduce the burden of manual assessment. The single Convolutional Neural Network (CNN) algorithm recognizes the handwritten characters but its accuracy is suppressed when dataset volume is increased. In proposed work stacking and soft voting ensemble mechanisms that address multiple CNN models to recognize the handwritten characters. The performance of the ensemble mechanism is significantly better than the single CNN model. This proposed work ensemble VGG16, Alexnet and LeNet-5 as base classifiers using stacking and soft voting ensemble approaches. The overall accuracy of the proposed work is 98.66% when the soft voting ensemble has three CNN classifiers.

1. INTRODUCTION

OCR gives a significant contribution to the digitization of documents using various semantic machine learning and deep learning algorithms. Many educational organizations, business processes, banking systems and health care systems collect the information in documents which include important information like medical reports, cheques, legal contracts, students answer sheets and assignments and receipts. These documents take more place to store so it is necessary to find the solution to transform document content in digital formats. The OCR is the best method to convert documents in machine editable format. The most major element of any system for managing paperless documents is optical character recognition. Utilizing the OCR technology, you can search for specific keywords within printed text. Similar to revising any other text document, you can make adjustments to the scanned document.

Shortly, recognition of handwritten text incorporates a key role in virtual assistants, online assessment of handwritten assignments, license plate recognition, psychometric test, etc. With the COVID-19 era, online learning has become the norm. In this situation, educational institutions are using hybrid models for exams. viz. classes online but tests by traditional methods such as online proctoring but handwritten answers. In this situation, the need for handwriting recognition has become acute. For languages like Marathi, it is an absolute need due to a lack of prior solutions.

Marathi offline handwritten character recognition is a very complicated task due to the adaptation of a person's writing manner, writing strokes, and writing posture with a person's moods or behavior during writing. The style of handwriting varies from person to person, so whatever handwritten images are added to the training and testing datasets should be more

in number and cover different styles of the handwriting of different persons. Real-time applications like handwritten cheque bank processing, accessing the address on postal letters, evaluating students' handwritten answer sheets, and collecting handwritten text from various applications are utilized for the digitization system using OCR implementation in various regional languages. The digitization of handwritten documents is handled by various approaches like semantic machine learning algorithms and deep learning algorithms by specifying different parameters to improve accuracy.

In a variety of computer vision research areas, single classifiers like Support vector machines (SVM), Artificial Neural Networks (ANN), and CNN are implemented for recognition but these classical algorithms give poor performance for imbalance datasets. The performance of the existing algorithm is reduced as the dataset class values are imbalanced and the computational load is not handled by these algorithms. There are a lot of challenges to train a dataset using a single estimator. The single estimator is more subtle to the contributed input for the trained feature data, so there are more chances of high variance. Also when making predictions, the model highly depends on an insufficient number of features. An ensemble can perform superior to a single contributing model and generate better estimates. An ensemble eliminates the spread or divergence of the model's performance and predictions. The ensemble learning approach is employed in many current research problems for COVID-19 detection using chest X-ray images [1] and for Monaural Speech Separation.

To address these issues, we propose an ensemble mechanism that groups VGGNet, AlexNet, and LeNet-5 CNN models to recognize Marathi's handwritten characters.

The ensemble mechanism employed two novel approaches

namely voting and stacking classifiers and the performance metrics of the proposed work are analyzed. The burden of high computational dataset load is handled by an ensemble mechanism that splits the training dataset over multiple base classifiers instead of a single classifier and the performance of the ensemble mechanism improves as the number of base classifiers increases in the ensemble mechanism. Ensemble mechanisms using various CNN models implemented and analyzed the recognition of characters and it is very appropriate to utilize various models and datasets rather than just one CNN model. The significance of the proposed ensemble learning approaches is to track increasing divergence among the various CNN models and datasets in view to overcome the issue of the over-fitting on the instances of the training dataset using cross validation in stacking classifiers. In this paper LeNet, VGG16, and LeNet5 CNN models are engaged together to recognize Marathi handwritten characters with a dataset using ensemble approaches.

2. RELATED WORK

Handwriting recognition has already been implemented by various researchers for Indian regional languages. The significant work was implemented and analyzed for Malayalam and Telugu, Devanagari, Kannada, and Gurmukhi languages using Wavelet Transform, Regional Zone with Structural Features, CNN, Fuzzy zoning, Haar wavelet transform with SVM, Fuzzy Adaptive Resonance Theory Map, and transfer learning deep learning algorithm. Marathi is part of the Devanagari script which contains a very complex structure of compound characters which are different from other languages. So, the Marathi language needs separate attention for handwritten characters and words using CNN, SVM, Artificial neural network, and KNN algorithm. These algorithms significantly recognize the Marathi handwritten characters but they misclassify some of the handwritten images also there is an occurrence of over-fitting with these algorithms so in the proposed work to cover misclassified images and to overcome overfitting issues we ensemble CNN models using stacking and voting approaches.

The Marathi language is part of the Devanagari script which contains 12 classes of vowels, 36 classes of consonants, and the remaining 10 classes of digits. The SVM, KNN, decision tree, and random forest machine learning algorithms are being used for Indian languages.

To achieve the approach of an ensemble learning method, we first need to select our base models to be aggregated. There are two major approaches to ensemble mechanism: heterogeneous and homogeneous learners from different classifier models and the same types of classifier multiple models respectively. This paper proposed and implemented heterogeneous CNN ensemble models like stacking and voting classifiers by combining various CNN models like VGG16/19, Alexnet, and Lenet-5.

The SVM, KNN, random forest, and Naïve Bayes are machine learning algorithms [2] implemented for digit recognition using WEKA with 90% highest accuracy. The external feature-based qualities were selected for the classification of tomato [2] using KNN and SVM. The features were reduced up to 30% to limit the computation complexity without effect on classification performance. The SVM and KNN with zoning features of Devanagari compound character [3] for recognition with approximately 96% accuracy. Marathi

handwritten digits and characters' features were extracted using the HOG method [4] and classified using SVM and KNN with efficient performance. The score of recognition of handwritten Gurumukhi characters and numerals [5] is 87.9% for 13,000 test samples using the random forest algorithm.

The deep learning CNN [6] algorithm extracted features of CT scan images to predict the Lung Carcinoma. The stacking ensemble Schemes [7] employed for Persian handwritten digit recognition along with deep learning algorithms. This softmax-based voting approach is better for two-model than three-models. The VGG 16 CNN architecture [8] was implemented for Devnagari characters' handwritten recognition for the small dataset of 58 classes with 95% accuracy. The Malayalam handwritten character recognition [9] was implemented using Hidden Markov Model and ANN classifier. Handwritten text recognition and categorization of handwritten names of cities and using various deep learning models [10] which include CNN layer, RNN layers, and CTC model. The Kazakh and Russian languages datasets were used for this task. The softmax-based voting with any number of models [11] compared with majority voting [11] with an odd number of models to improve the accuracy. The group ensemble [12] approach in which all ensemble models are interconnected to a single ConvNet instead trained separately in multiple ConvNets to reduce the computation cost. Four block multi-block deep convolutional neural networks (CNN) [13] use two pretrained VGG16 and ResNet models which are utilized separately for extracting the features of images and these features are ensemble through three models like SVM, random forest, and linear regression to improve recognition accuracy performance. Image classification using Multi-Layers Deep Features Fusion and selection-based technique contributed significantly by fusion of VGG and Inception V3 deep learning feature set and selecting a sturdy feature set using Multi Logistic Regression controlled Entropy-Variations method [14]. The selected significant features utilized for final classification purposes affect both image classification accuracy and computational time. If the image resolution is poor then this method is unable to fetch significant features of images.

Multilingual handwritten numeral recognition using various CNN architecture without a fusion approach [15] employed recognition as 10 numeral classes. This model was deployed for eight regional languages with satisfactory performance. This model gives the best accuracy for the English language but for the Marathi language, its performance is poor due to the writing style of the Marathi language. Ensemble classifier approach [16], DCNN extracted the features of images and these feature sets pass through three SVM classifiers using different forms of the kernel like linear, Gaussian, and Polynomial. The three classifier scores are concatenated using ensemble mechanism models like majority voting, Vote for strongest decision, and Vote for sum decision. Handwritten Digit Recognition [17, 18] using Hybrid CNN-SVM Classifier employed by feeding CNN extracted features of handwritten digit images to SVM for further classification. Ensemble Learning Approach for Handwritten Digit Recognition achieves good accuracy by decreasing the complexity of the model for a small-scale dataset of digits. The dataset is distributed to multiple CNN base models randomly and ensembles all classifier's output using the bagging average model approach. The stacked ensemble neural network is used for handwritten Marathi numerals recognition. The stacked ensemble concatenates the pre-trained base learners to create

a meta-learning classifier that outputs the final target labels. DeepNet Devanagari model used CNN model for classification of basic characters of ancient Devanagari documents categorized into 33 classes. The accuracy of 93% was achieved after 30 epochs for 5484 images. Farsi handwritten digits are recognized using a pragmatic convolutional bagging ensemble learning [19]. This presented model integrates CNN as base classifiers like VGG16, Xception, ResNet18, and bagging weighted majority ensemble learning for Farsi digit recognition.

3. METHODOLOGY

3.1 CNN architecture

In the proposed ensemble learning framework the CNN classifiers are incorporated for prediction. The SVM, KNN and Artificial neural network (ANN) are traditional machine learning algorithms which are utilized for image classification but these algorithms explicitly extract features of images [20, 21] using Histogram of gradient and Gabor filter methods. CNN is primarily a better option for image recognition because CNN itself is a feature extractor so there is no need for explicit feature extraction using classical image feature extraction algorithms. CNN learned the features retrieved using CNN are low level features as compared to traditional image processing methods so prediction accuracy is better than SVM, ANN and KNN algorithms.

The CNN model is the base model for the recognition of images with good accuracy. The working of CNN [22] flows with different layers with their significance. Every layer in CNN is responsible for image preprocessing tasks like image feature extraction and classification. The CNN contains the Convolution layer, Activation Layer, Pooling Layer, and Classification layer. The input image is fed to the convolution layer which involves the image. The high and low levels of the image feature map are produced by the convolution layer and fed to the activation layer. The activation and deactivation of neurons are processed by the Relu activation function. The activation layer generates rectified feature maps and passes through the pooling layer which down samples the dimension of the feature map. The final Feature map is generated by hidden layers of CNN and fed for classification through fully

connected or Dense output layers. The last layer of the output layer fired the Softmax activation function which categorized the images into different classes with the highest probability score. The flow of the CNN system is shown in Figure 1.

The Softmax equation as shown below in Eq. (1).

$$P(a^{(i)}) = \phi_{softmax}(a^{(i)}) = \frac{e^{a^{(i)}}}{\sum_{j=0}^c e^{a_k}} \quad (1)$$

where,

$$a = w_0z_0 + w_1z_1 + w_2z_2 + \dots + w_mz_m = \sum_{k=0}^m w_kz_k$$

where, (z) feature map value of one training sample instance is, w is weightage of ith sample instance. The softmax function formulates the probability of training instance z⁽ⁱ⁾ of respective class c. The softmax function gives the probability for each class c=1.....,m. The probabilities of each instance concerning all classes sum up to one.

In this paper, Marathi handwritten character recognition was implemented using LeNet, VGG Net, and AlexNet.

LeNet-5 [23] is 7 layered CNN architecture that trains 60k trained parameters and is used for grayscale handwritten digits for classifying digits in 10 classes. LeNet-5 was developed by LeCun et al. in 1998 and presented their work in the paper "Gradient-Based Learning Applied to Document Recognition". The AlexNet [24] network is very similar to the LeNet-5 network but deeper and has around 60 million parameters trained by AlexNet. AlexNet proposed by Alex Krizhevsky, Geoffrey Hinton and Ilya Sutskever in ImageNet Classification with Deep Convolutional Neural Networks". AlexNet is composed of 8 layers with new additional features like max-pooling and relu activation function. VGG Net [25] was designed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition". VGG utilized smaller filters than AlexNet and attached to 16/19 layers. The stacking of smaller filters is better for improving accuracy and has a smaller number of parameters. The number of layers employed in each CNN model along with a number of parameters is shown in following Table 1.

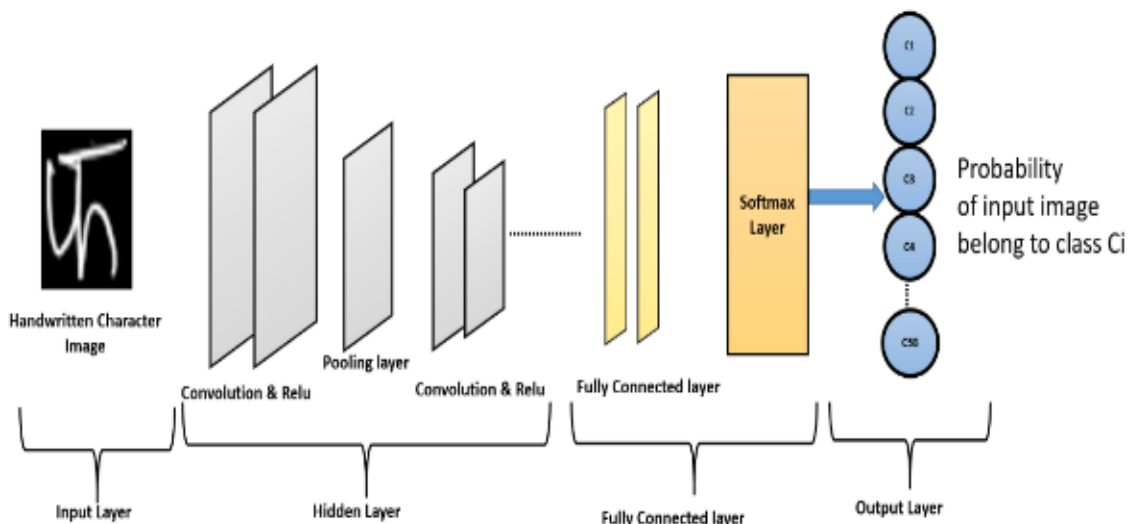


Figure 1. Marathi handwritten character recognition using CNN model

Table 1. CNN architectures

CNN Architecture	No. of Layers	Error Rate	Number of Parameters
LeNet-5	5	--	60 thousand
AlexNet	8	15.3%	60 million
VGG	16/19	7.3%	138 million

The proposed model ensemble framework is integrated with Lenet-5, VGG-16 and Alexnet CNN classifiers which are more reliable for image recognition with their own capabilities. The computation power of VGG-16 is more as it works with more parameters in more layers as compared to AlexNet and LeNet-5 architectures, so VGG-16 CNN model is incorporated with AlexNet and LeNet-5 to reduce the burden of high computation and also avoid overfitting due to training of more features.

3.2 Ensemble learning model

Ensemble Learning [26] is the mechanism of combining multiple classifier models to solve recognition or classification tasks. In the proposed paper the various ensemble approach like voting and stacking classifier is utilized for Marathi handwritten character recognition by assembling multiple CNN architectures to improve the classification accuracy.

3.2.1 Voting ensemble learning model

The voting ensemble method is usually employed for the classification task. When a single approach model exhibits bias to a specific aspect, the voting classification algorithm is a system for the detection and can be a good alternative. Using this strategy, a generalized fit of each individual model can be obtained. The Voting classification algorithm focuses on the performance of numerous classifiers thus substantial inaccuracies or wrong classifications from any one estimator won't impact the outcomes. Strong performance from other classifiers can make up for a bad performance from one of them.

In this method, multiple base classifiers are worked to make predictions for each image instance of the training dataset. The prediction outcome of each base model is the votes of the respective model by each model which is generated by an ensemble of all base models. The majority of base models'

outcomes prediction is the final prediction of the ensemble model. The voting ensemble method is practically employed using a Voting classifier which is a casing of a set of different models that are trained and predicted in parallel to utilize the different specialties of each base model.

In the proposed work soft voting is preferred because of the SoftMax classifier and categorical cross entropy is employed for classification. Soft voting deals with prediction in terms of class probabilistic information. Soft voting predicts the category with the highest category probability score, averaged over all the individual base classifiers. The target class will be the highest probability score. The architecture of the soft voting ensemble classifier is shown in Figure 2.

The Soft voting as shown in Eq. (2)

$$Y_{pred} = \frac{1}{N_{BC}} \sum_{BC} (Y_{pred1}, Y_{pred2}, Y_{pred3}) \quad (2)$$

3.2.2 Stacking ensemble learning model

In stacking, estimations from contributing ensemble members are combined in the most effective way possible using a machine learning model. The heterogeneous weak classifier employed in stacking classifiers in level-1 and incorporating them by learning a meta-classifier to predict the class of image relied on weak classifier's predictions. The weak classifiers were trained on various subsets of training data that were resampled utilizing cross validation methods in the suggested stacking ensemble strategy. So the stacking ensemble is more reliable for classification as the classifier employed on various sample dataset and its outcomes are stacked to the meta estimator to improve the result of prediction.

The stacking ensemble learning model [27] runs through multiple base classifier models at level one on the same training dataset. The Stacking ensemble learning approach avail concept of cross validation where training dataset is distributed in to n-folds and followed consecutively in n sequence, in n-1 folds are employed to train level one base classifier; in each sequence level-1 base classifiers are claimed to remaining portion of dataset that was not part of model training in each loop.

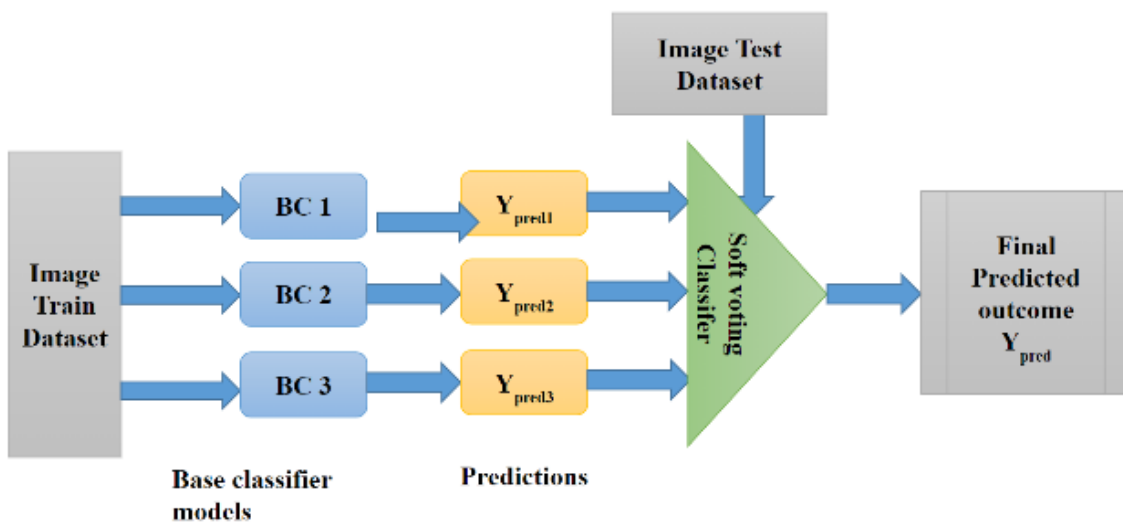


Figure 2. Soft voting ensemble learning model

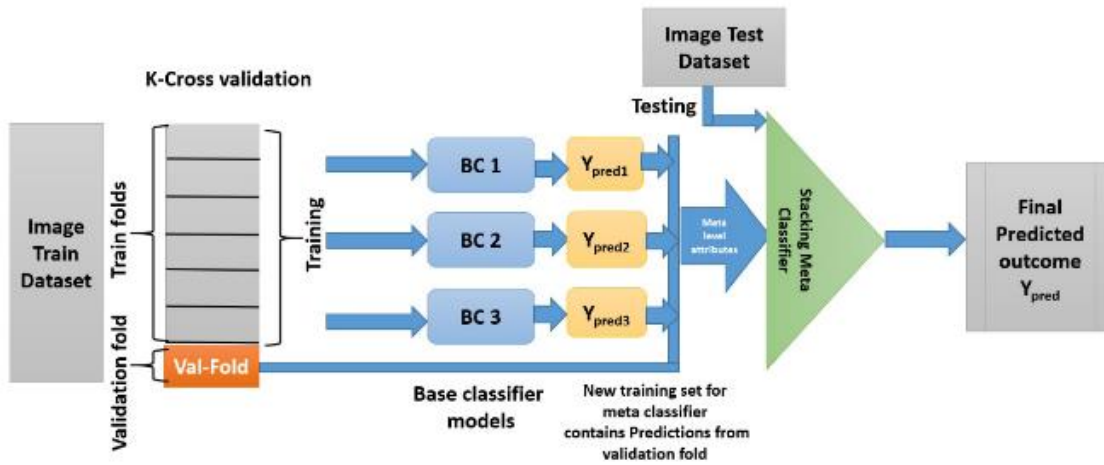


Figure 3. Stacking ensemble learning model

Base classifiers learn the training dataset side by side and its predictions are assembled. The predictions of these base classifier models are fed as input to a second-level stacking classifier which is a stacking meta-classifier or final estimator. The final estimator is trained on a new train dataset generated by all predictions of base classifiers.

In this way, the stacking classifier is employed in parallel and concatenated predictions in the level-1 passed over meta classifier at level-2 for training and test the test dataset over it for prediction based on various weak base classifiers.

The cross-validation is employed on base models and finally, predictions are aggregated on the meta-classifier as shown in Figure 3. Stacking the ensemble method with cross-validation results in better accuracy and overcomes the overfitting problem.

The base classifiers estimate probability distribution over all possible class scores. So, the prediction of base classifier at level-1 BCs trained over training dataset z is probability distribution as shown in Eq. (3):

$$P^{BC}(z) = (P^{BC}(c_1|z), P^{BC}(c_2|z), \dots, P^{BC}(c_n|z)) \quad (3)$$

where, $\{c_1, c_2, \dots, c_n\}$ are classes, and $P^{BC}(c_i|z)$ is the probability that dataset instance z to class c_i as predicted by the base classifier at the respective level. The dataset instance z belongs to class c_i as predicted by the base classifier BC. The class/category c_j with maximum class probability ($P^{BC}(c_j|z)$) is estimated by classifier BC. The probabilities of estimation of each class by each base classifier model are represented by Meta level attributes i.e. ($P^{BC_j}(c_i|z)$) where, $i=1, \dots, n$ and $j=1 \dots N$.

4. METHODOLOGY IMPLEMENTATION

4.1 Dataset and pre-processing

The input images fed to image preprocessing [28] before fed to prediction models. Marathi handwritten dataset handled for the ensemble learning experiment contains a total of 58 classes distributed as 120000 for training and 20000 for testing purposes. The training image instances are marked as stated by their respective classes while the test dataset is managed

without a label. The 58 classes of dataset contain a total of 12 classes of vowels, 36 classes of consonants, and the remaining 10 classes of digits. Data augmentation is a mechanism employed to increase training data instances by applying specific methods like random vertical and horizontal flip and shift, random rotation, random brightness, and contrast. All images were reshaped to a 32*32 scale and transformed into a grayscale channel.

4.2 Observation and discussion of experimental result

The main objective of the paper is to achieve more accuracy in solving the Marathi handwritten character recognition problem. As mentioned before, LeNet-5, AlexNet, and VGG are ensembled for Marathi handwritten character recognition with 5 epochs for the same training dataset. In the proposed work various ensemble mechanism approaches were employed for better accuracy like voting and stacking with different base classifiers like LeNet-5, AlexNet, and VGG. The ensemble mechanism employed with various numbers of CNN base classifier combinations over training dataset and prediction performance is analyzed based on handwritten image samples from the test dataset. The performance of the proposed stacking and voting ensemble learning approach with 2 and 3 base classifiers compared with single classifier prediction. The base CNN classifiers with various combinations that offered higher accuracy will be considered as the final prediction. The accuracy of the proposed model is assessed from the prediction of various ensemble mechanism approaches with test image dataset instances. The highest test accuracy model will take the final determination.

The training and testing accuracy of recognition of handwritten characters using a single CNN classifier along with the number of parameters involved in each CNN model shown in Table 2.

Table 2. Single base classifier accuracy and parameters

Base CNN Classifier	Parameters	Accuracy train	Accuracy test	Training Time in Second
Lenet-5	871,166	95	94	107.8
AlexNet	348,718	96	95.12	204
VGG 16	33,784,558	97.15	95	563.319

The following Figure 4 depicts the performance of prediction using a single base classifier like LeNet-5, AlexNet, and VGG16.

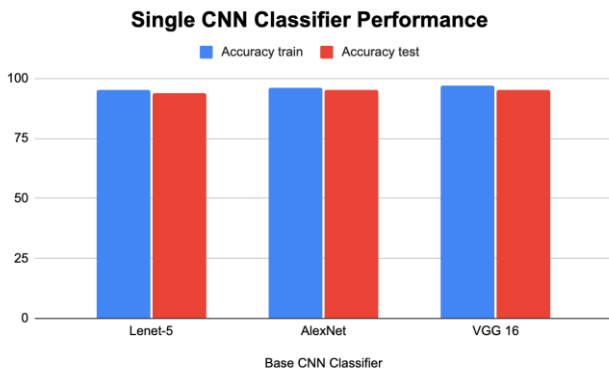


Figure 4. Graphical visualization of performances of single CNN classifier

The test and train accuracy for all two and three base classifiers with all CNN architectures using the soft voting ensemble approach is visualized and analyzed in Figure 5.

The experimental observation of the performance of the ensemble learning mechanisms using multiple base classifiers with various combinations is shown in Table 3.

As test and train accuracy shown in Table 3, it is observed that the training computation time is more when three CNN classifiers are integrated using a soft voting ensemble method as three models will take more time for prediction based on highest category probability score and its training and testing accuracy is better than the stacking model employed using two CNN classifiers as three classifiers are contributing for prediction. The stacking ensemble model incorporated using three CNN classifiers takes more time for training than stacking ensemble models with two classifiers.

Table 4 demonstrates the F1 Score evaluation metrics for vowels, digits, and some consonants when soft voting

ensembles for three base classifiers. As per the evaluation metric of F1-score, it is observed that some characters look like other characters with less score.

The characters, अः and अः are looking the same so their score is less due to the same appearance of characters.

The overall performance of the stacking classifier employed using two or three base classifiers is not satisfactory due to insufficient data of all classes to train further and not flexible in view of time required for its execution in level-1 and level-2. The visualization of the performance of the stacking ensemble learning approach is shown in Figure 6.

Based on the argmax of the sums of the predicted probability of the different estimators that make up the ensemble, the soft voting predicts the class label. The final class will be determined by the one with the highest weighted and summed probability after the Soft Voting Classifier takes into account the probabilities offered by each CNN classifier. As the experimental result of the soft voting classifier visualized in Figure 5, it is observed that all involved CNN models.

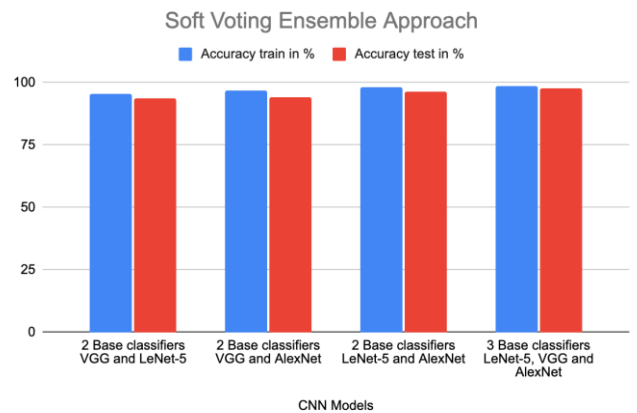


Figure 5. Soft voting ensemble approach using VGG, LeNet, and AlexNet CNN base classifiers

Table 3. Overall accuracy of proposed ensemble mechanism approaches

Ensemble Learning	CNN Models Base classifier with various combinations	Accuracy train in %	Accuracy test in %	Training Time in seconds
Soft Voting	2 Base Classifiers VGG and LeNet-5	95.34	93.55	502.39
	2 Base Classifiers VGG and AlexNet	96.6	94.02	647.48
	2 Base Classifiers LeNet-5 and AlexNet	98.12	96.33	185.31
	3 Base Classifiers LeNet-5, VGG and AlexNet	98.66	97.84	1517.7
Stacking	2 Base Classifiers VGG and LeNet-5	95.05	94.02	2400.4
	2 Base Classifiers VGG and AlexNet	96.09	96.01	2995.1
	2 Base Classifiers LeNet-5 and AlexNet	97.98	97.01	794.48
	3 Base Classifiers LeNet-5, VGG, and AlexNet	97.67	96.09	3312.3

Table 4. The F1 score for vowels, digits, and some consonants when soft voting ensemble for three base classifiers

Vowels	अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ	अं	अः
F1 Score in %	96	98	97	98	98	98	96	97	94	98	97	94
Sample Consonants	क	ख	ग	घ	ङ	च	छ	ज	झ	-	-	-
F1 Score in %	96	98	98	99	96	96	99	99	95			
Digits	१	२	३	४	५	६	७	८	९	०	-	-
F1 Score in %	98	96	99	97	98	99	95	99	99	98		

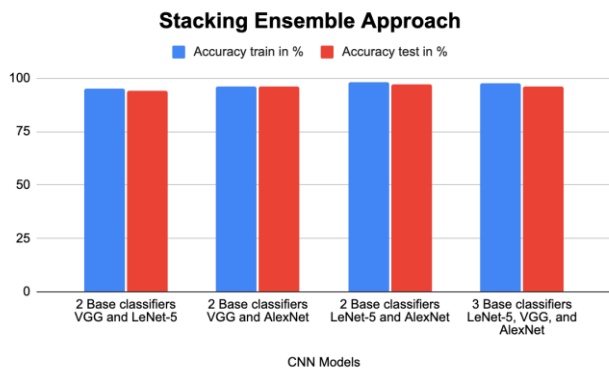


Figure 6. Stacking ensemble approach using VGG, LeNet, and AlexNet CNN base classifiers

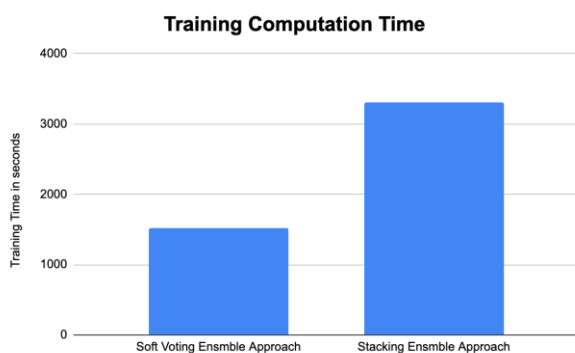


Figure 7. The training time is taken by soft voting and stacking ensemble learning approaches using 3 base classifiers

The stacking ensemble model is employed in a stacked manner where prediction outcomes of all CNN base classifiers provide the meta classifier and predict the final result. In stacking meta classifiers learners operate on the prediction result of base classifiers in previous level so the final prediction results in recognition of images. The results of stacking classifiers which are employed using various combinations of CNN classifiers are analyzed in Figure 6. The performance of stacking and soft voting ensemble approaches for the test dataset are good and more than 96%, but soft voting ensemble approach works better than stacking ensemble approach. When Soft voting and stacking ensemble models are employed with three CNN classifiers, the performance of soft voting model is better than stacking model as training computation time of the soft voting model's is also less than stacking model's training computation time. The stacking classifier is employed in two levels of classification; it will take more computation time and so accuracy also affects the performance. Based experimental results shown in Figure 5 and Figure 6 of soft voting and stacking ensemble model, interpreted as both approaches require more training computation time when any base classifier combined with VGG CNN classifier because VGG employed more parameters and more layers.

Figure 7 depict the training time required for soft voting and stacking ensemble approaches using 3 base classifiers.

5. CONCLUSIONS

CNN architectures like Lenet-5, VGG16 and AlexNet were

utilized individually for recognition of handwritten characters with sufficient training and testing accuracy. However, in this paper, ensemble learning approaches are proposed and compared with performance of a single classifier. The ensemble model combines the predictions of two or three CNN base classifiers for Marathi handwritten recognition and is fulfilled with better training and testing accuracy than single base classifier recognition. In this paper, soft voting, and stacking heterogeneous ensemble learning approaches are operated over the handwritten character image dataset. The performance of soft voting and stacking ensemble approaches deployed with Lenet-5, VGG16 and AlexNet CNN base classifiers with various combinations of two or three base classifiers are better than single base classifiers. The soft voting and stacking ensemble learning approaches are deployed on the same dataset in different manners, the soft voting give the better performance of handwritten character recognition than stacking ensemble approach in which the level 2 meta classifier recognize the image based on meta-data generated by all base classifiers's prediction instead of whole dataset. The training time required for the soft voting ensemble approach is less than the stacking approach because the stacking model incorporates two levels. The future work will extend the recognition of Marathi handwritten characters using Adaboost and Xgboost ensemble approach.

REFERENCES

- [1] Balasubramaniam, S., Kumar, K.S. (2023). Optimal Ensemble learning model for COVID-19 detection using chest X-ray images. *Biomedical Signal Processing and Control*, 81: 104392. <https://doi.org/10.1016/j.bspc.2022.104392>
- [2] Shweta, S.D., Barve, S.S. (2021). External feature based quality evaluation of Tomato using K-means clustering and support vector classification. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, pp. 192-200. <https://doi.org/10.1109/ICCMC51019.2021.9418420>
- [3] Deore, S.P. (2021). Devanagari handwritten compound character recognition using various machine learning algorithms. In *2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4)*, Bangalore, India, pp. 1-6. <https://doi.org/10.1109/C2I454156.2021.9689425>
- [4] Chikmurge, D., Shriram, R. (2019). 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
- [5] Kumar, M., Jindal, M.K., Sharma, R.K., Jindal, S.R. (2020). Performance evaluation of classifiers for the recognition of offline handwritten Gurmukhi characters and numerals: A study. *Artificial Intelligence Review*, 53(3): 2075-2097. <https://doi.org/10.1007/s10462-019-09727-2>
- [6] Yadav, A., Badre, R. (2020). Lung carcinoma detection techniques: A survey. In *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, Bhimtal, India, pp.

- 63-69.
<https://doi.org/10.1109/CICN49253.2020.9242633>
- [7] Haghghi, F., & Omranpour, H. (2021). Stacking ensemble model of deep learning and its application to Persian/Arabic handwritten digits recognition. *Knowledge-Based Systems*, 220, 106940.
- [8] Deore, S.P., Pravin, A. (2020). Devanagari handwritten character recognition using fine-tuned deep convolutional neural network on trivial dataset. *Sādhanā*, 45(1): 1-13. <https://doi.org/10.1007/s12046-020-01484-1>
- [9] Gurav, Y., Bhagat, P., Jadhav, R., Sinha, S. (2020). Devanagari handwritten character recognition using convolutional neural networks. In 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, pp. 1-6.
<https://doi.org/10.1109/ICECCE49384.2020.9179193>
- [10] Nurseitov, D., Bostanbekov, K., Kanatov, M., Alimova, A., Abdallah, A., Abdimanap, G. (2021). Classification of handwritten names of cities and handwritten text recognition using various deep learning models. arXiv preprint [arXiv:2102.04816](https://doi.org/10.48550/arXiv.2102.04816).
<https://doi.org/10.48550/arXiv.2102.04816>
- [11] Jhang, K. (2020). Voting and ensemble schemes based on CNN models for photo-based gender prediction. *Journal of Information Processing Systems*, 16(4), 809-819.
- [12] Chen, H., Shrivastava, A. (2020). Group ensemble: Learning an ensemble of convnets in a single convnet. arXiv preprint [arXiv:2007.00649](https://doi.org/10.48550/arXiv.2007.00649).
<https://doi.org/10.48550/arXiv.2007.00649>
- [13] Chirra, V.R.R., Uyyala, S.R., Kolli, V.K.K. (2021). Virtual facial expression recognition using deep CNN with ensemble learning. *Journal of Ambient Intelligence and Humanized Computing*, 12(12): 10581-10599. <https://doi.org/10.1007/s12652-020-02866-3>
- [14] Rashid, M., Khan, M.A., Alhaisoni, M., Wang, S.H., Naqvi, S.R., Rehman, A., Saba, T. (2020). A sustainable deep learning framework for object recognition using multi-layers deep features fusion and selection. *Sustainability*, 12(12): 5037. <https://doi.org/10.3390/su12125037>
- [15] Gupta, D., Bag, S. (2021). CNN-based multilingual handwritten numeral recognition: A fusion-free approach. *Expert Systems with Applications*, 165: 113784. <https://doi.org/10.1016/j.eswa.2020.113784>
- [16] Gupta, J.D., Samanta, S., Chanda, B. (2018). Ensemble classifier-based off-line handwritten word recognition system in holistic approach. *IET Image Processing*, 12(8): 1467-1474. <https://doi.org/10.1049/iet-ipr.2017.0745>
- [17] Ahlawat, S., Choudhary, A. (2020). Hybrid CNN-SVM classifier for handwritten digit recognition. *Procedia Computer Science*, 167: 2554-2560. <https://doi.org/10.1016/j.procs.2020.03.309>
- [18] Bhujade, R.K., Pandit, A. (2018). A Survey on various methodologies of automatic handwritten character recognition using neural network. *International Journal of Emerging Technology and Advanced Technology*, 8(2): 328-332.
- [19] Nanekaran, Y.A., Chen, J., Salimi, S., Zhang, D. (2021). A pragmatic convolutional bagging ensemble learning for recognition of Farsi handwritten digits. *The Journal of Supercomputing*, 77(11): 13474-13493. <https://doi.org/10.1007/s11227-021-03822-4>
- [20] Jinturkar, A.A., Khanale, P.B. (2020). Segmentation & recognition of marathi handwritten numeral strings using multiple features. *International Journal of Emerging Technology and Advanced Technology*, 10(5): 2250-2459.
- [21] Rajpoot, V., Dubey, R., Mannepalli, P.K., Kalyani, P., Maheshwari, S., Dixit, A., Saxena, A. (2022). Mango plant disease detection system using hybrid BBHE and CNN approach. *Traitement du Signal*, 39(3): 1071-1078. <https://doi.org/10.18280/ts.390334>
- [22] Narang, S.R., Kumar, M., Jindal, M.K. (2021). DeepNetDevanagari: A deep learning model for Devanagari ancient character recognition. *Multimedia Tools and Applications*, 80(13): 20671-20686. <https://doi.org/10.1007/s11042-021-10775-6>
- [23] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278-2324. <https://doi.org/10.1109/5.726791>
- [24] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.
- [25] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://doi.org/10.48550/arXiv.1409.1556).
<https://doi.org/10.48550/arXiv.1409.1556>
- [26] Nandan, K.V.P., Panda, M., Veni, S. (2020). Handwritten digit recognition using ensemble learning. In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 1008-1013. <https://doi.org/10.1109/ICCES48766.2020.9137933>
- [27] Acharya, S., Pant, A.K., Gyawali, P.K. (2015). Deep learning based large scale handwritten devanagari character recognition. In Proceedings of the 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Kathmandu, Nepal, pp. 1-6. <https://doi.org/10.1109/SKIMA.2015.7400041>
- [28] Rajpoot, V., Dubey, R., Khan, S.S., Maheshwari, S., Dixit, A., Deo, A., Doohan, N.V. (2022). Orchard Boumans algorithm and MRF approach based on full threshold segmentation for dental X-Ray images. *Traitement du Signal*, 39(2): 737-742. <https://doi.org/10.18280/ts.390239>