A new approachto image classification based on adeep multiclass AdaBoosting ensemble

Hasan Asil[1], Jamshid Bagherzadeh[2]
Department of Computer Engineering, Faculty of Electrical and Computer Engineering, Urmia University, Iran

Article Info

Article history:

Received Apr 20, 2019
Revised Mar 11, 2020
Accepted Mar 30, 2020

Keywords:

Boosting
Convolution
Deep learning
Ensemble methods
Image classification

ABSTRACT

In recent years, deep learning methods have been developed in order to solve the problems. These methods were effective in solving complex problems. Convolution is one of the learning methods. This method is applied in classifying and processing of images as well. Hybrid methods are another multi-component machine learning method. These methods are categorized into independent and dependent types. Ada-Boosting algorithm is one of these methods. Today, the classification of images has many applications. So far, several algorithms have been presented for binary and multi-class classification. Most of the above-mentioned methods have a high dependence on the data. The present study intends to use a combination of deep learning methods and associated hybrid methods to classify the images. It is presumed that this method is able to reduce the error rate in images classification. The proposed algorithm consists of the Ada-Boosting hybrid method and bi-layer convolutional learning method. The proposed method was analyzed after it was implemented on a multi-class Mnist data set and displayed the result of the error rate reduction. The results of this study indicate that the error rate of the proposed method is less than Ada-Boosting and convolution methods. Also, the network has more stability compared to the other methods.

Corresponding Author:

Jamshid Bagherzadeh,
Department of Computer Engineering,
Faculty of Electrical and Computer Engineering,
Urmia University,
Urmia, Iran.
Email: j.bagherzadeh@urmia.ac.ir

1. INTRODUCTION

In recent years, deep learning methods have gained a lot of attention and were used in many types of research. In deep learning, first, the nonlinear features of several layers are extracted [1]. Then, they are transferred to a classifier, and finally, they are sent to a combiner layer, in order to perform the combination and prediction. The more hierarchy of layers is (deeper), the more nonlinear features are obtained and better results are presented. These methods are used to solve various complex problems [2]. One of these networks is the convolution method, which has many applications in image processing [3]. This network was originally designed to process images, and various convolution-based models have been developed to classify the images.

On the other hand, hybrid methods are common machine learning techniques. These methods combine different predictions to provide better and more accurate problem-solving results [4]. These methods are used for solving classification problems and have applied in different classification problems in order to provide various predictions [5].

No algorithms are optimal in all areas. Every learning algorithm is limited to a specific model. In fact, if data assumptions are not true, errors will occur [6]. If the model parameters are defined for a training set, the model might not be appropriate for some data, for which another model can produce better results [2]. Ensemble classifiers are among the multicomponent classifiers defined to produce better results than a single-component classifier [7]. In such classifications, ensemble classifiers are used to obtain better results. Ensemble methods vary in how they create different classifiers and how they combine basic classifiers with respect to their weights [8].

In fact, there are two ensemble frameworks: dependent (serial) and independent (parallel) [9]. In a dependent framework, the output of a classifier is used in the next classifier. Therefore, the knowledge of previous iterations can be used to direct learning in next iterations [10]. Boosting is an instance of such frameworks. In the second framework (known as independent), every classifier is created independently. Their outputs are combined with those of voting methods [11].

So far, many techniques have been proposed for classifying the images. Some of these classifications are bi-class and some others are multi-class [8]. It should be noted that multi-class classifications have many applications. Creating a robust model in line with network stability, reducing classification error and decreasing data dependency in the model are some of image classification issues [12].

The present study proposes a new model based on deep learning methods and hybrid method to improve and reduce the error rate of multi-class image classification. On the other hand, combining these techniques results in higher stability, compared to previous methods. This hybrid boosting method is a bi-layer convolution related hybrid method. As you know, the goal of boosting is to strengthen and improve poor learning, which attempts to strengthen its learning and make the model stronger. In this study, the bi-layer convolution technique was considered a weak learning method.

2. BACKGROUND

In the past, deep learning networks were used to solve different problems such as image classification, object recognition, image extraction, etc. [13].These networks employed various methods for optimization. Convolutional neural networks (CNNs) are among the most important deep learning networks possessing several layers and serving as a very powerful method for machine vision optimization. These networks consist of three layers named convolutional, pooling, and fully connected. CNNs have been used for image classification.

A deep learning model was introduced by Daniel Frits et al. for image classification. This model combines a deep network with PixelRNN and DCGRAN models for image recognition. These models were implemented in PixelRNN and DCGRAN for handwritten data [14]. Gan Project was carried out for image classification, too. It is used in a k-class categorization [15]. Another classification project included the auxiliary deep generative models (ADGMs). ADGMs consist of various sets of encoders and decoders. This project was executed on the MINIST dataset. According to the results, AtlasRBF algorithm produced 1.5 times more convergence than deep generative models, virtual adversarial, and ladder methods [16-18].

A major approach to deep networks is to create different layers for learning the features [19, 20]. An instance of such networks is an eight-layered network designed by Krizhevsky et al. [21]. It is a deep network architecture based on the convolution of a deep eight-layered network. Given the success of this network, it is used widely to solve many problems such as video classification [22], facial recognition, and action detection [20]. Other relevant projects include the use of a convolution a weak learner in the ensemble AdaBoosting algorithm [23].Weak learner is a learner that no matter what the distribution over the training data is will always do better than chance, when it tries to label the data. This means that the learner algorithm is always going to learn something, not always completely accurate. The combination of a convolutional network and AdaBoosting method has also been used in numeration. The aim of this paper was to propose a model based on the combination of AdaBoosting method and a two-layered convolutional network to produce better results.

3. RESEARCH METHOD

In the proposed approach, ensemble methods and deep leaning were employed to improve image classification. For this purpose, a convolutional neural network was combined with the AdaBoosting method. Every ensemble problem consists of four sections: a training set, a basic learner, a driver, and a composer [10]. Likewise, the proposed method includes the following section:

a. Training set:A training set contains labelled samples used for training. In this study, the MINIST standard dataset was employed for image classification.
b. Basic learner: A basic learner is a learning algorithm used for learning a training set. In this project, the basic learner was used to implement deep learning. For this purpose, a convolutional neural network was used as a learning algorithm. However, the CNN consisted two convolutional functions in every

classifier for more convergence. On the other hand, the convolutional functions are based on error rate reduction in learning.

c. Driver: This component is used to create different classifiers.
d. Composer: This component is used to combine classification methods. In this project, the boosting technique was used.

Figure 1 shows a schematic view of the proposed method for combination.

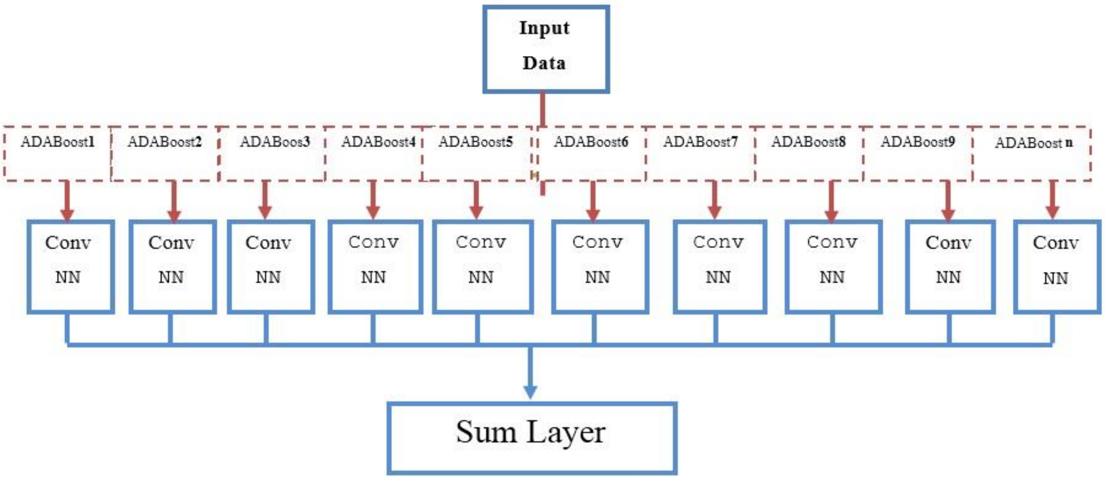


Figure 1. The general architecture of the proposed method

The AdaBoosting network was created and combined with the convolutional neural network to obtain a powerful learner from weak convolutional learners. Accordingly, the convolutional network consists of three layers: convolutional, pooling, and fully connected. The fully connected layer is responsible for computing the classification [23, 24]. In the proposed algorithm, two deep convolutional layers of every class are responsible for weak learning. The multiclass boosting network is defined in the following way [13].

$$F: X \rightarrow \{1 \ldots M\} \tag{1}$$

In this structure,X is all data in Class M also every xirefers to a class number of a label number of an M-classification. F(X) is trained globally and developed in the following way [9].

$$F(x) = \arg\max <Yk , F(X)>k=1...M \tag{2}$$

In (2), YK¬ indicates the label of a k-classification.In fact, the output of F(X) is as follows [13].

$$F(x) = [f1(x)... fM(x)] \in RM \tag{3}$$

On the other hand, the multiclass boosting method divides the learning process into several subclasses to improve learning. Therefore, gi: X→RD shows training functions.

$$F(X)= \sum \propto_t g_t(x) \tag{4}$$

In this function, t and α show the frequency and correlation coefficient in each iteration, respectively. In this method, GD-MCBoost was employed to improve learning [13]. Considering the learning classification, the proposed algorithm used a minimization policy in which a descending gradient was employed [15]. Therefore, the minimum weak learner is selected in the following way:

$$g=\arg\min R[f;g] \tag{5}$$

The (6) indicates the function used for minimization [7].

$$\delta R[f; g] = -\frac{1}{2||D||} + \sum_{Xi,Zi\in D} < g(x_i)w(x_i) > \quad (6)$$

The following algorithm is a part of the source code of the proposed method:

Algorithm 1. The algorithm structure

```
cnn.layers = {
    struct('type', 'i') %input layer
    struct('type', 'c', 'outputmaps', 6, 'kernelsize', 5) %convolution layer
    struct('type', 's', 'scale', 2) %sub sampling layer
    struct('type', 'c', 'outputmaps', 12, 'kernelsize', 5) %convolution layer
    struct('type', 's', 'scale', 2) %sub sampling layer
};


Sample code for any classification%
opts.alpha = 1;
opts.batchsize = 50;
opts.numepochs = 500;
cnn,er5]= cnntrain(cnn, train_x5, train_y5, opts);
D(1,5), bad4] = cnntest(cnn, test_x5, test_y5);
cnn = cnnsetup(cnn, train_x6, train_y6);
 opts.alpha = 1;
opts.batchsize = 50;
opts.numepochs = 500;
[cnn,er6] = cnntrain(cnn, train_x6, train_y6, opts);
[D(1,6), bad5] = cnntest(cnn, test_x6, test_y6);
cnn = cnnsetup(cnn, train_x7, train_y7);
opts.alpha = 1;
opts.batchsize = 50;
opts.numepochs = 500;
[cnn,er7] = cnntrain(cnn, train_x7, train_y7, opts);
[D(1,7), bad6] = cnntest(cnn, test_x7, test_y7);
cnn = cnnsetup(cnn, train_x8, train_y8);
cnn,er10]= cnntrain(cnn, train_x10, train_y10, opts);
D(1,10), bad9] = cnntest(cnn, test_x10, test_y10);
Increase the result by min%
er=cat(1,er1,er2,er3,er4,er5,er6,er7,er8,er9,er10);
er=min(er);
```

4. RESULTS AND ANALYSIS

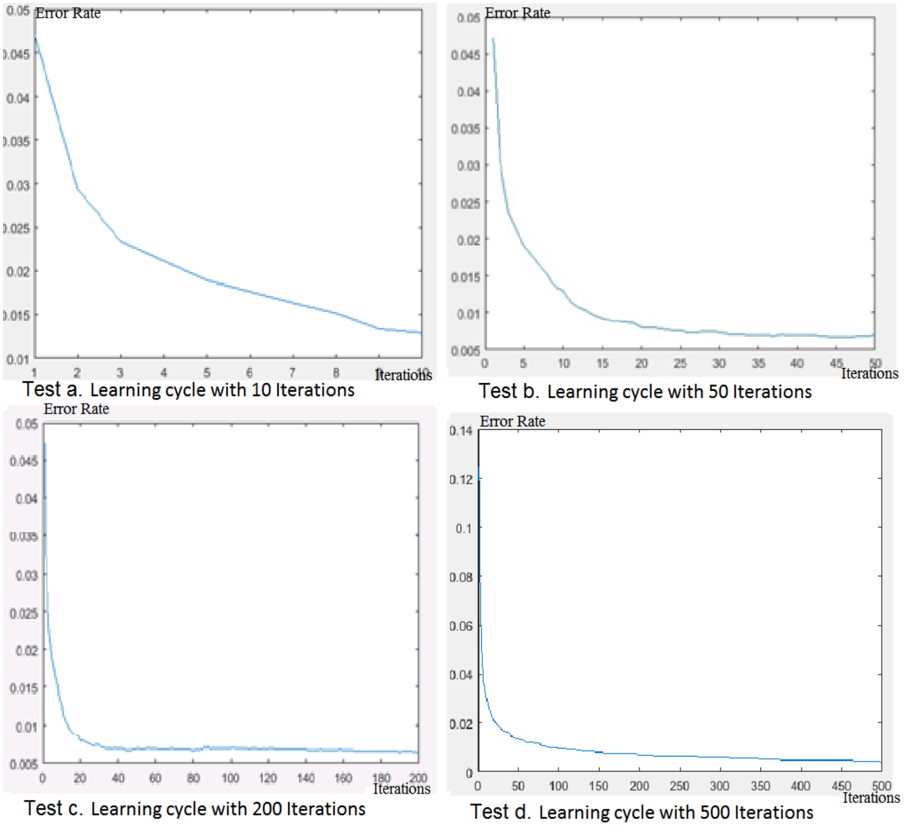
The proposed algorithm was implemented on the MINIST dataset in MATLAB for evaluation [25]. The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems [26]. The database is also widely used for training and testing in the field of machine learning [27, 28]. It was created by "re-mixing" the samples from NIST's original datasets. Furthermore, the black and white images from NIST were normalized to fit into a 28x28 pixel bounding box and anti-aliased, which introduced grayscale levels. The MNIST database contains 60,000 training images and 10,000 testing images [29]. Section 4.1 to 4.4 shows the testing results.In these sections,the CNN, AdaBoosting method, and the proposed ensemble multiclass convolutional AdaBoosting method were evaluated to compare the results. On the other hand, every test was evaluated with different iterations to analyze the results better.

4.1. AdaBoost

Figure 2 indicates the process of learning the AdaBoosting algorithm for different iterations. Accordingly, the error rate decreased when the number of iterations increased. In this algorithm, the sample space was divided into classes. Then the learning process was carried out on those classes in the AdaBoosting method. Naturally, the number of iterations and length of learning increased. In other words, there were 100, 500, and 2000 processes in 10, 50, and 200 iterations.The learning cycle find the set of parameters to optimize the error function. In the end, the model and parameters are with the smallest error.

4.2. Single-layered CNN

A CNN consists of convolutional, pooling, and fully connected layers [30]. Therefore, a comparison should be made between this method and the proposed method. The CNN was evaluated on the MINIST dataset. Figure 3 shows the results of the CNN. Figure 2 indicates the CNN learning process in 10, 50, and 200 iterations. Accordingly, the CNN is slower than the AdaBoosting method.



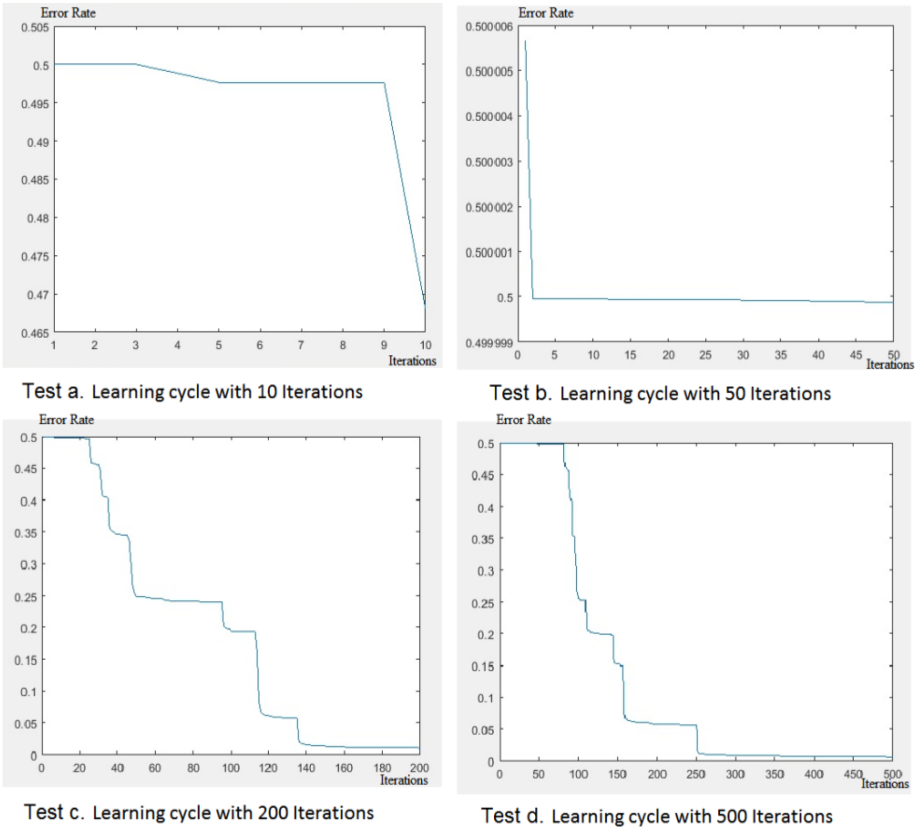
Test a. Learning cycle with 10 Iterations

Test b. Learning cycle with 50 Iterations

Test c. Learning cycle with 200 Iterations

Test d. Learning cycle with 500 Iterations

Figure 2.Error rate comparison based on the number of iterations in the in AdaBoosting



Test a. Learning cycle with 10 Iterations

Test b. Learning cycle with 50 Iterations

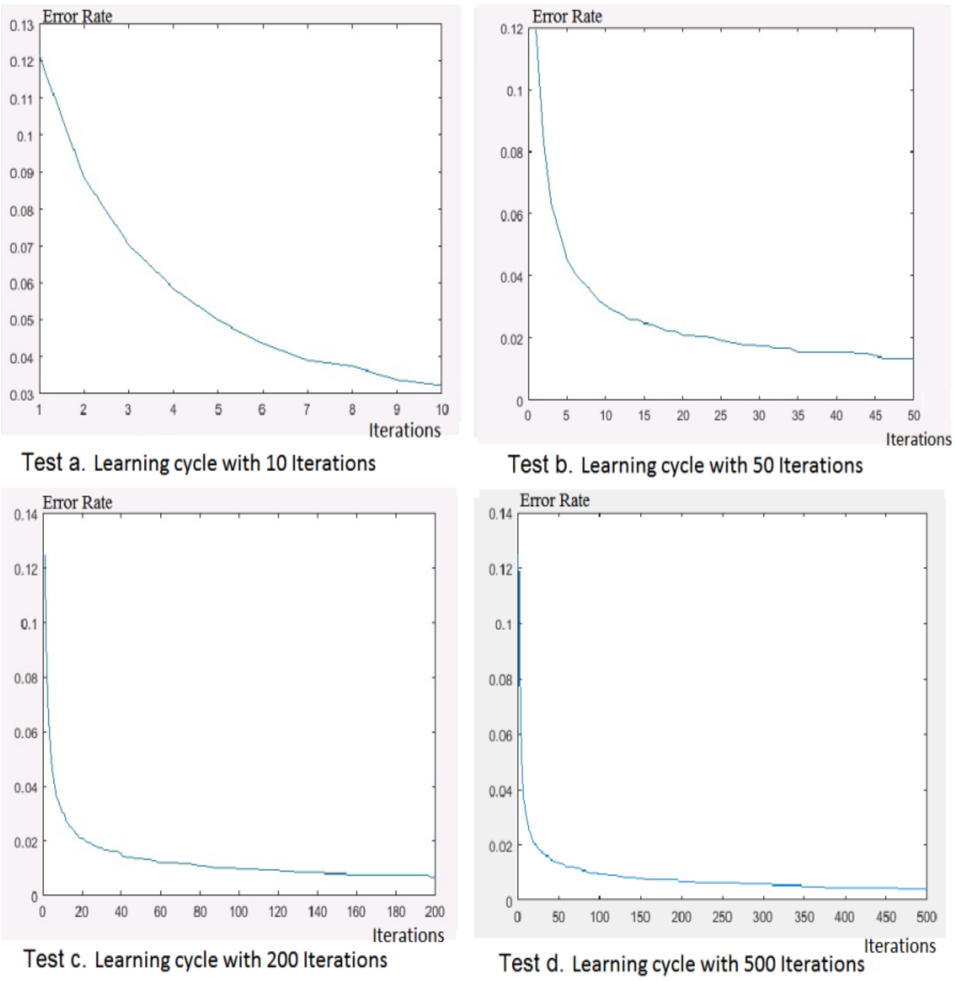
Test c. Learning cycle with 200 Iterations

Test d. Learning cycle with 500 Iterations

Figure 3. Error rate comparison based on the number of iterations in the CNN

4.3. Two-layered CNN

Considering the use of two-layered CNNs in weak learning, a convolutional deep learning network was formed. The classification results of that network were evaluated to produce better results. Figure 4 indicates the results of that network.



Test a. Learning cycle with 10 Iterations

Test b. Learning cycle with 50 Iterations

Test c. Learning cycle with 200 Iterations

Test d. Learning cycle with 500 Iterations

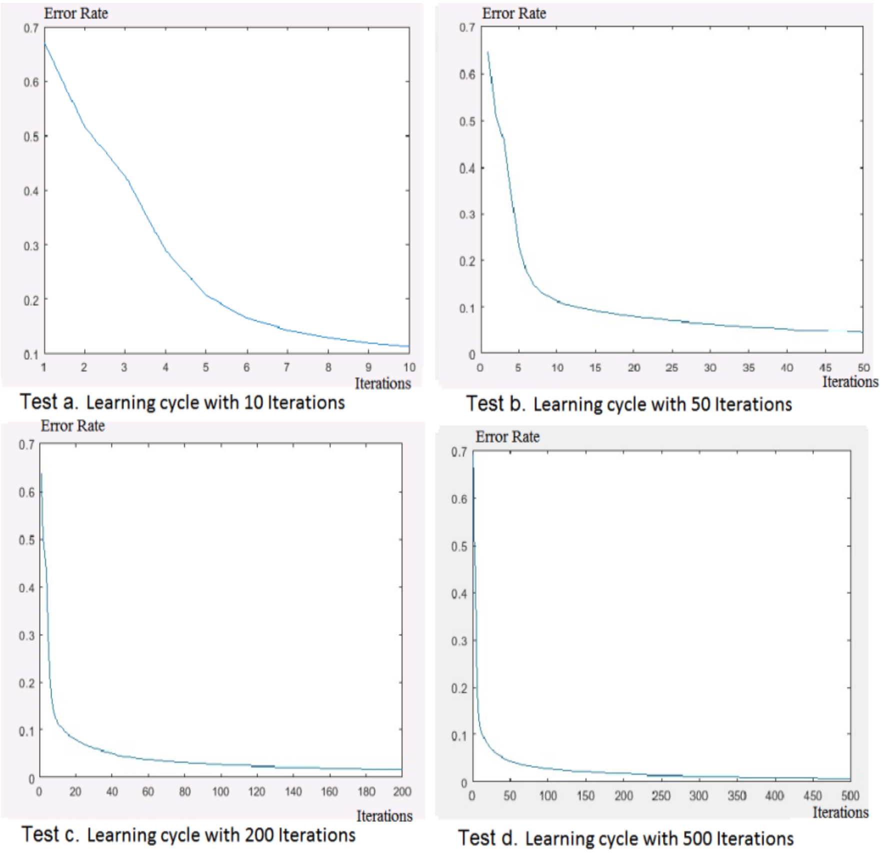
Figure 4. Error rate comparison based on the number of iterations in the two-layered CNN

4.4. Deep multiclass AdaBoosting

The proposed method was anensemble of AdaBoosting and convolutional methods. The sample space was divided into ten units in combination with the convolutional method. Figure 5 shows the results of learning process in different iterations.

4.5. Results and discussion

In order to evaluate the proposed method, the image classification was conducted on the Mnist data set with different algorithms. The methods used for implementing these experiments were presented in sections 4.1 to 4.4. The mono- and bi-layer convolution methods were carried out to evaluate the AdaBoosting methods. The error rates were determined based on the number of repetitions. The results of these experiments are shown in Table 1. For example, in 500 repetitions, the error rate of Ada-boosting, bilater convolution, and deep Ada-boosting are 0.008, 0.0089, and 0.0067 respectively. On the other hand, if the number of repetitions increased then a better convergence can be achieved in the proposed method.According to the results, the image classification error rate in the proposed method was 8% better than the previous method and had better stability as well. Table 1 shows the error rates in three different situations. Figure 6 shows a comparison drawn between the proposed algorithm and other methods.



Test a. Learning cycle with 10 Iterations

Test b. Learning cycle with 50 Iterations

Test c. Learning cycle with 200 Iterations

Test d. Learning cycle with 500 Iterations

Figure 5. Error rate comparison based on the number of iterations in deep multiclass AdaBoosting

Table 1. Error rate comparison based on algorithms and iterations

Row	Iterations	AdaBoosting	Single-Layered CNN	Two-Layered CNN	Convolutional Multiclass AdaBoosting
1	10	0.0135	0.4680	0.0323	0.4501
2	50	0.0094	0.5000	0.0037	0.0446
3	200	0.009	0.0197	0.0089	0.0168
4	500	0.008	0.087	0.0081	0.0067

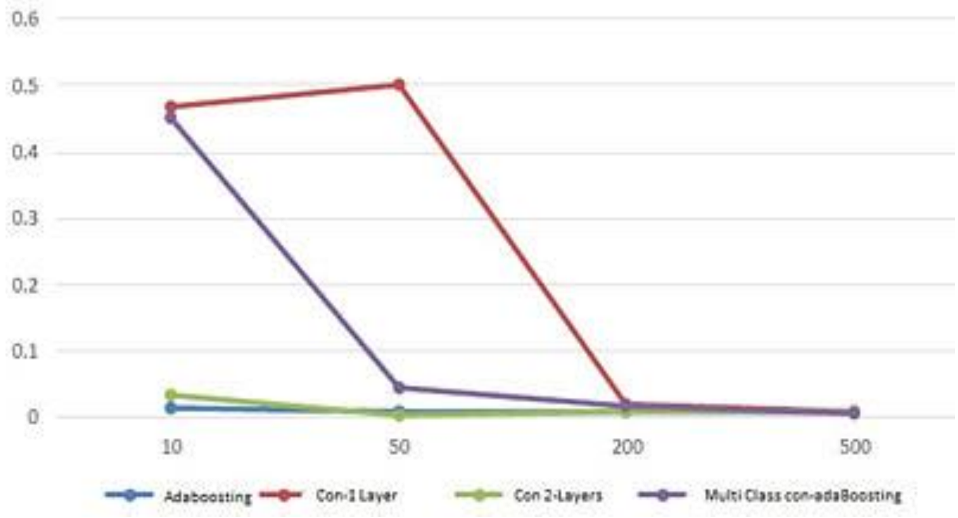


Figure 6. The error rate of algorithms based on iterations

5. CONCLUSION

Deep learning is now used for solving various complex problems. Convolution is one of the deep learning methods, which is used for classifying images. The hybrid methods are considered as a multi-component learning method that intends to strengthen the learning process by using combined approaches. Boosting is one of these methods. This study seeks to find a method for optimizing and reducing errors when classifying images. So far, various methods and models have been presented for this classification. The proposed method is based on a combination of boosting method and deep convolution technique. Boosting has a set of basic learning that improves and strengthens the poor learning in each step. In the proposed method, the bi-layer convolution network is considered as basic learning. This method was implemented using MATLAB software and evaluation was carried out using handwritten data. In this algorithm, the Mnist data set was applied. After evaluating the results of the proposed method, it was analyzed using convolution and boosting methods. The results indicated that in 500 repetitions, the error rate was decreased by 8%. On the other hand, since the boosting method was used, the proposed model has higher stability.In the future, it is possible to improve the proposed algorithm by strengthening the basic algorithm and creating some connections in the weighting process.

REFERENCES

[1] H. F. Pardede, *et al*., "Convolutional Neural Network and Feature Transformation for Distant Speech Recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, pp. 5381-5388, 2018.

[2] S. Khan and Kirubanand V.B., "Comparing machine learning and ensemble learning in the field of football," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 4321-4325, 2019.

[3] R.Caruana and A. N. Mizil, "An Empirical Comparison of Supervised Learning Algorithms," *ICML '06 Proceedings of the 23rd international conference on Machine learning, USA*, pp. 161-168, 2006.

[4] E. Alpaydin, "Introduction to Machine Learning," Massachusetts, USA, *MIT Press*, 2004.

[5] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEEtransaction on neural networks*, vol. 13, no. 2, pp. 415-425, 2002.

[6] T. T. Khuat and M. H. Le, "Ensemble learning for software fault prediction problem with imbalanced data," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 4, pp. 3241-3246, 2019.

[7] M. Ghosh and Prabu P., "Empirical Analysis of Ensemble Methods for the Classification of Robocalls in Telecommunications," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 4, pp. 3108-3114, 2019.

[8] S. Ravidas and M. A. Ansari, "Deep learning for pose-invariant face detection in unconstrained environment," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 1, pp. 577-584, 2019.

[9] J. Tanha, "Ensemble approaches to semi-supervised learning," *Thesis*, University of Amsterdam, 2013.

[10] A. Rasmus, *et al*., "Semi-Supervised Learning with Ladder Networks, Neural and Evolutionary Computing," *arXiv:1507.02672*, 2015.

[11] J. Tanha, *et al*., "An AdaBoost Algorithm for Multixlass Semi-supervised Learning," *IEEE 12th International Conference on Data Mining (ICDM)*, vol. 1, pp. 1116-1121, 2012.

[12] M. S. H. Al-Tamimi, "Combining convolutional neural networks and slantlet transform for an effective image retrieval scheme," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 4382-4395,2019.

[13] M. Moghimi, *et al*., "Boosted Convolutional Neural Networks," *Proceedings of the 27th British Machine Vision Conference*, pp. 24.1-24.13, 2016.

[14] A. vanOord, *et al*., "Pixel Recurrent Neural Networks," *Proceedings of the International Conference on Machine Learning*, vol. 48, pp. 1747-1756, 2016.

[15] A. Oden, "Semi-Supervised Learning with Generative Adversarial Networks," *Data Efficient Machine Learning workshop at ICML 2016*, 2016.

[16] H. Valpola, "From neural PCA to deep unsupervised learning," in *Advances in Independent Component Analysis and Learning Machines*, pp. 143-171, 2015.

[17] L. Maaløe, *et al*., "Auxiliary Deep Generative Models," *Proceedings of the 33rd International Conference on Machine Learning*, vol. 48, pp. 1445-1453, 2016.

[18] R. Jensen and Q.Shen, "New Approaches to Fuzzy-Rough Feature Selection," *IEEE Transactions on FuzzySystems*, vol. 17, no. 4, pp. 824-838, 2009.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large scale image recognition," *arXiv: 1409.1556*, 2014.

[20] C. Szegedy, *et al*., "Going deeper with convolutions," *arXiv: 1409.4842*, 2014.

[21] A. Krizhevsky, *et al*., "Image net classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097-1105, 2012.

[22] A. Karpathy, *et al*., "Large-scale video classification with convolutional neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1725-1732, 2014.

[23] J. Zhu, *et al*., "Multi-class adaboost," *Statistics and Its Interface*, vol. 2, pp. 349-360, 2009.

[24] N. Karianakis, *et al*., "Boosting convolutional features for robust object proposals," *arXiv: 1503.06350*, 2015.

[25] M. J. Saberian and N. Vasconcelos, “Multiclass Boosting: Theory and Algorithms,” *Advances in Neural Information Processing System*, 2011.

[26] Qiao, Yu, “The MNIST Database of handwritten digits,” Retrieved 18 August 2013.

[27] J. C. Platt, “Using analytic QP and sparseness to speed training of support vector machines,” *Advances in Neural Information Processing Systems*, pp. 557-563, 1999.

[28] T. A. Assegie and P. S. Nair, “Handwritten digits recognition with decision tree classification: a machine learning approach,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 4446-4451, 2019.

[29] Y. LeCun, *et al*., “MNIST handwritten digit database,” [Online]. Available: http://yann.lecun.com/exdb/mnist/.

[30] K. Kumar N., *et al*., “An ensemble multi-model technique for predicting chronic kidney disease,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 2, pp. 1321-1326, 2019.

BIOGRAPHIES OF AUTHORS

Hasan Asil is a lecturer in the Faculty computer of Islamic Azad University, Azarshahr Branch. He is also PHD Student at Urmia University.His research Fields are Query Optimization in Database, Software Development Methodologies, Data Mining, Machine Learning.

Jamshid B. Mohasefi received his B.S. in computer engineering from Sharif University of Technology in Iran at 1996 and M.S. in computer engineering from Tarbiat Modares University in Iran at 1999. He got his PhD in computer engineering from Indian Institute of Technology Delhi (IITD) in India at 2006. He is currently associate professor and dean of computer and electrical engineering faculty at Urmia University, Iran. He is also head of Computer Emergency Response Team (CERT) center at Urmia University, Iran. His current research interests include network security, machine learning, and internet of things.