



A pragmatic convolutional bagging ensemble learning for recognition of Farsi handwritten digits

Y. A. Nanekaran^{1,2} · Junde Chen¹ · Soheil Salimi³ · Defu Zhang¹

Accepted: 16 April 2021 / Published online: 26 April 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Recognition of handwritten digits is one of the most important and challenging issues in recent decades in the field of computer science. Its cursive nature, the right to left writing styles of words and characters as well as various digits shapes have imposed curiosity among numerous researchers to impose a lot of efforts on the recognition of handwritten Farsi numbers. In order to improve the recognition accuracy of Farsi handwritten digit recognition, the pragmatic CBWME network structure model based on convolution bagging weighted majority ensemble learning is developed by integrating the convolution neural network (CNN) and bagging weighted majority ensemble learning. For base classifiers, we applied the VGG16, ResNet18, and Xception architectures and explored the bagging weighted majority ensemble learning in combining the base classifiers results, which are later used in identifying Farsi handwritten digits. The performance of the CNN models (VGG16, ResNet18, and Xception) and CBWME model was evaluated by comparing their results. From the experimental result analysis, it was observed that the proposed CBWME model achieved the best average recognition accuracy (97.65%), followed by the Xception model (95.9%), ResNet18 model (93.75%), and VGG16 model (90.26%) in HODA dataset. The accuracy orders were the same as in IFHCDB and CENPARMI datasets. The CSE model attained the best result with rate of (99.876%) compared with the other studies.

Keywords Farsi digits handwritten recognition · Bagging strategy · Ensemble learning · Average accuracy · Xception · Deep neural networks

✉ Defu Zhang
lycmym@163.com

Extended author information available on the last page of the article

1 Introduction

Recognition of handwritten numbers is a very important research area in machine vision and optical character recognition (OCR) fields [1, 2]. The Farsi handwritten recognition (FHR) is a complicated task: due to the specific writing nature and styles of Farsi cultivars, it is much more difficult to distinguish Farsi handwritten cultivars than English cultivars. On the other hand, the application of Farsi digits handwritten recognition (FDHR) in various fields has led to the innovation of numerous ideas to expand and improve the quality of recognizing Farsi handwritten digits, which also called for the interest in this study [3–5].

The deep learning (DL) model can be transformed into simple and more complex features by simplifying features and can be used well to identify handwritten characters; thus, to improve the accuracy of recognition, many researches focus on constructing the models of handwriting digits recognition by using DL [6–9]. Farahbakhsh et al. [10] proposed a model for Farsi handwritten digits recognition based on AlexNet. Data augmentation technique was used to produce more data. HODA dataset was used to evaluate the proposed method. Considering the fact that researchers have made some changes in architecture model of AlexNet and high number of diversity of data, the resulted method showed better results. Latif et al. [11] proposed a model based on DNN architecture to recognize handwritten characters of some Eastern languages like Eastern Arabic, Farsi, Urdu, Devanagari and Western Arabic. Nanekaran et al. [12] used CNN for recognition of Farsi handwritten digits, which achieved the result of 99.45%. Akhlagi and Ghods [13] proposed a method to recognize handwritten phone numbers, which seek to use DNNs to get the best results. They first changed the digit strings into single digits and then tried to recognize them. Elkhayati et al. [14] introduced a new method based on CNNs and computational geometry algorithms (CG), which were used to recognize isolated handwritten. IFHCDB has been used in the proposed methods. Sahlol et al. [15] designed a mixed method for handwritten digits recognition, in which CENPARMI method was used. The method used different models of CNNs and compared their results. Safarzadeh and Jafarzadeh [16] proposed a method for recognition of Farsi handwritten digits in which CNN and recurrent neural network (RNN) were used. Also, in order to delete segmentation level used in usual methods, a connectionist temporal classification loss function was used. The method was efficient in Farsi handwritten digits' recognition. HODA dataset was used in this study. Modhej et al. [17] proposed a method for Farsi handwritten digits' recognition using brain's hippocampus dentate gyrus. The proposed method has solved this problem, and the accuracy of handwritten digits recognition has developed. Authors have also introduced excitation and inhibition steps, which play important roles in getting the results better. HODA dataset was used for recognizing and evaluating Farsi handwritten digits. Parseh et al. [18] introduced a model based on CNN, but the difference was that a nonlinear multi-class support vector machine classifier was used instead of a fully connected layer. Considering the changes in structure and architecture of CNN, the results have high accuracy, and the studies in this field have provided better results in comparison with other results.

However, the efficiency and performance accuracy of these models require a lot of training data, which in turn upgrades the computational complexity of the models [19–21]. In addition, the training datasets are usually limited, and these individual models may not obtain the correct distribution of the sample data from the hypothesis space, which may result in low performance accuracy of the models [22–25]. Thus, to overcome this limitation, ensemble learning methods were introduced to improve the performance accuracy [26–28]. To improve the recognition accuracy of Farsi handwritten digit recognition, our contributions are as follows: (1) the CBWME network structure model based on convolution bagging weighted majority ensemble learning is developed by integrating the convolution neural network and bagging weighted majority ensemble learning. For base classifiers, we applied the VGG16, ResNet18, and Xception architectures and explored the bagging weighted majority ensemble learning in combining the base classifiers results, which are later used in identifying Farsi handwritten digits; (2) the performance of the CNN models (VGG16, ResNet18, and Xception) and CBWME model was evaluated and analysed by comparing their results.

The organization of this paper is as follows: Section 2 presents the preliminaries of the data preparation, convolutional neural network models and the performance evaluation. The experimental results are given in Sect. 3. Finally, the discussion, hypothesis and limitations, and conclusions are presented in Sects. 4, 5, and 6, respectively.

2 Methodology

The study methodology is based on three steps, shown in Fig. 1: data preparation, model architecture, and performance evaluation. In data preparation, we described the datasets used in this paper. In model architecture, different models of CNN were amalgamated to form the model of CBWME. In performance evaluation, the performance of the models can be evaluated by proposed formulas.

2.1 Data preparation

There are several databases of handwritten Farsi digits available. In this study, three different datasets were used. The first database named HODA, which was collected from approximately 12,000 entrance examination registration forms. The dataset contained around 102,352 samples, which included Farsi digit examples. The next group dataset of this paper is Isolated Farsi Handwritten Character Database (IFHCDB). IFHCDB has been used in this study, which includes grayscale images, which had 17,740 samples. The last dataset is called CENPARMI, which includes 11,000 instances.

2.2 Convolutional neural network models

2.2.1 VGG16

VGG16 is composed of 16 layers and accepts image inputs of 224×224 RGB size and passes them through a stack of convolutional layers with the fixed filter size of

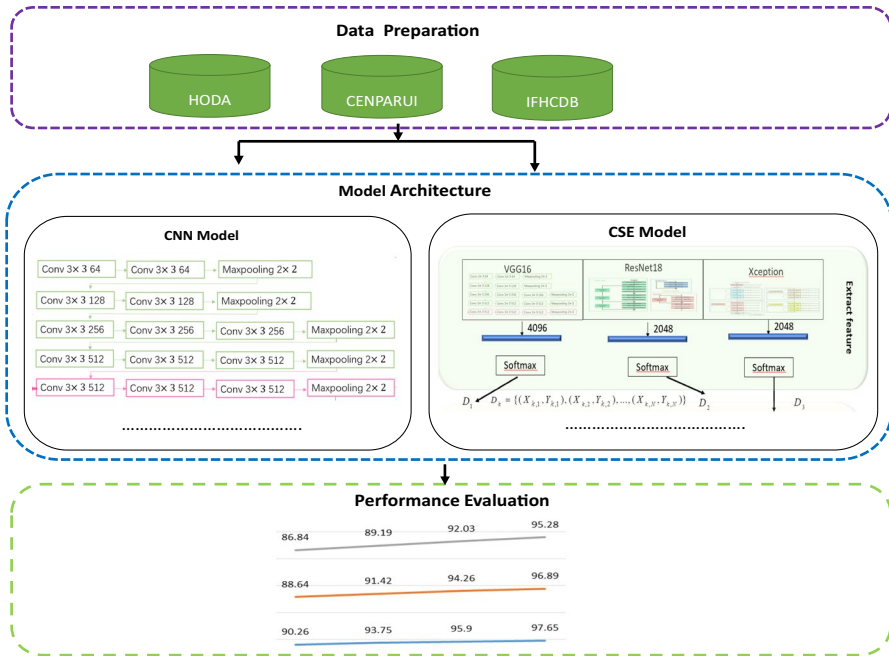


Fig. 1 Illustration of the overall flowchart

3 × 3 and the stride of 1. In this architecture, there are 5 max pooling filters encapsulated between convolutional layers for the purpose of down-sampling the input representation [29]. Following the stack of convolutional layers are the 3 fully connected layers, which consists of 4096, 4096 and 1000 channels, respectively. And finally the last layer in the architecture is the soft-max layer [30]:

2.2.2 ResNet18

There are four convolutional blocks in network configuration, and related convolutional layers have 3, 5, 7 and 3 layers, respectively, and the size of convolutional kernel of each module is all 3 × 3. There are different numbers of majors in each convolutional blocks (from the figure, the numbers of residual modules in each convolutional block are 1, 2, 3, and 1, respectively). In network, pooling layer disappears in each convolutional block. Convolutional layer down samples pooling layer using two strides [31, 32].

2.2.3 Xception

This architecture is made of 36 convolutional layer to increase the extraction ability to classify the image, which creates 14 modules and punctuates residual connections except first and last modules. According to required size and channels (299 × 299 × 3), input image uses entry flow in the first module with 2 convolutional

layers allocating 32 and 64 filters and 3×3 kernel and modules 2–4, with kernel size of 3×3 and separable convolutional filters 128, 256 and 728 [33, 34]. In this section, map feature of $19 \times 19 \times 728$ is produced and passes the rings (modules fifth to twelfth) 8 times in middle flow by separable convolutional filters 728. Then, map feature is transferred from middle flow to front to final part, in which thirtieth module has used two separable calculation filter of 728 and 1024. Adding global average pooling and a fully connected layer before logistic regression as the last layer, two separable convolutional filters size of 1536 and 2048 are applied in final modules [33, 35, 36].

2.2.4 Model architecture of CBWME

The outline of the model architecture of Convolutional Bagging Weighted Majority Ensembles (CBWME) is shown in Fig. 2, in which it includes base classifier and meta classifier. For base classifier, probability of each predicted class of each CNN algorithm can be obtained. For meta classifier, we adapted the idea of bagging ensemble learning and obtained the better classifier by combing weighted majority of each base classifier.

The CNN layers including the input layer, convolution layers, pooling layers, fully connected layers and output layer, where the base classifiers generated are chosen in CBWME model. Our proposed ensemble classifier with weighted majority is far different from the bagging ensemble approach, which equally integrate the base classifier predictions by using an averaging model. The distinctness among the ensemble classifiers guarantees their uniqueness in performance during Farsi handwritten digits classification. Despite that, certain base classifiers may exist, which can be used to classify Farsi handwritten digits patterns and could be assorted to expand the distinctness as well as enhance the performance accuracy.

Likewise, certain base classifiers tend to show lower capability in determining some Farsi handwritten digits; thus, their influence in FHR pattern is reduced. To combine the results from three base classifiers, in proportion to their approximated performance

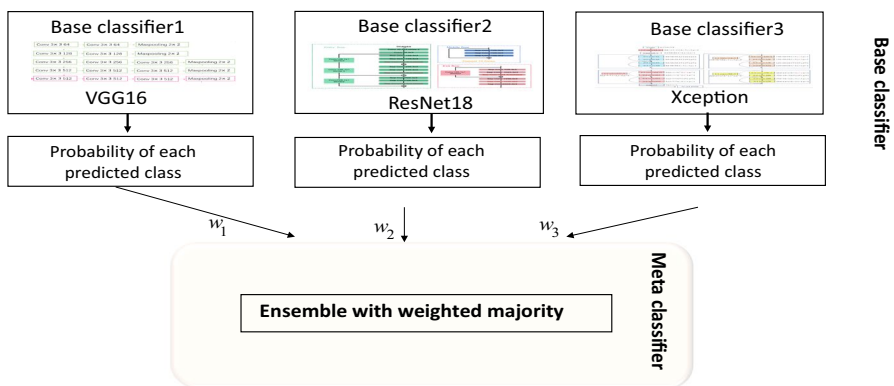


Fig. 2 The outline of the model architecture of CBWME

we used the weighted majority function, which facilitate the contribution of multiple classifiers in Farsi handwritten digits recognition pattern, which is as follows:

$$C_i = \arg \max \left(\sum_{k=1}^M w_k p_k(y|X_i) \right) \quad (1)$$

where X_i and y represent the i th input image and the vector of classified label, respectively. Taking an assumption that, there are five recognition classes of Farsi handwritten digits, the first class is (1, 0, 0, 0, 0). The M parameter denotes the number of base classifiers that are considered in the ensemble model. The probability of $p_k(y|X_i)$ represents the output value of k th base classifier, which is computed from the soft-max function in the output layer of k th base classifier. The weight w_k signifies a vector of weight for each Farsi handwritten digit class, which is identified based on the fraction of the total Farsi handwritten digits that were extracted. During training, the weights from the ensemble classifier validation dataset tend to be more robust and overcome the over-fitting problem.

2.3 CBWME algorithm

Based on the above analysis, the CBWME algorithm can be summarized as follows:

Input: image

Step1: Construct base classifier and obtain probability of each predicted class.

Step2: Ensemble with weighted majority.

Output: the result of classification

In step1, base classifiers can be constructed by VGG16, Res18, and Xception, which receive image and generate the probability of each predicted class. In step2, using weighted majority function decides the final result of each predicted class.

2.4 Performance evaluation

In this present study, some useful statistical metrics have been used to evaluate the classification models, namely accuracy, sensitivity, and specificity, which are calculated as follows:

$$\text{Accuracy} = \frac{ES + ER}{ES + ER + AS + AR} \quad (2)$$

$$\text{Sensitivity} = \frac{ES}{ES + AR} \quad (3)$$

$$\text{Specificity} = \frac{ER}{AS + ER} \quad (4)$$

where the values denoted by ES , ER , AS and AR are defined as follows. ES The number of samples that are correctly identified belongs to a particular class. ER The number of samples that are correctly rejected belongs to a particular class. AS The number of samples that are mistakenly identified belongs to a particular class. AR The number of samples that are mistakenly rejected belongs to a particular class.

3 Results and discussion

3.1 Experiment settings

Cross-validation is one of the different techniques for validating the same model in order to evaluate in what way the outcomes of a statistical analysis will generalize to a set of independent data. It is implemented where the goal is prediction, and one attempts to calculate how accurately a predictive model will do practically. Testing model's ability is considered as the main objectives of using cross-validation to predict new data that was not used in estimating it, through which one can overcome the problems like over-fitting or selection bias and to provide a viewpoint on how the model will be generalized to an independent dataset (i.e. an unknown dataset, for instance from a real problem).

One step in cross-validation involves adding a sample of data into complementary subsets, implementing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). In order to minimize the variability, in plenty of methods multiple rounds of cross-validation are done by means of various partitions, and the validation results are combined (e.g. averaged) over the rounds to give an evaluation of the model's predictive performance.

In k -fold cross-validation, the initial sample is randomly selected and entered into k equal-sized subsamples. Of the k subsamples, a single subsample is maintained as the validation data for evaluating the model, and the remaining $k - 1$ subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples implemented exactly once as the validation data. The k results can then be averaged to produce a single estimation. The positive points of this method comparing with repeated random subsampling are that both training and validation are implemented in observations, and each of them is used for validation exactly one time. It is common to use fivefold or tenfold cross-validation; however, k is generally an unfixed parameter. The amount of K has been chosen 5 in the proposed method.

In fact, we divided each dataset (HODA, IFHCDB, and CENPARMI) into 5 categories, respectively. During the experiment, in the first step, to obtain result a_1 the first set is used as test set and the other 4 as training sets. In the second step, to obtain a_2 , the second set is considered as a test set and the other 4 as training sets. The above processes are repeated until all sets have participated in test and training.

Table 1 The datasets and partition of training and test sets

Datasets	Process	Test set	Training sets	Results
HODA (102,352 samples)	Step 1	1	2 3 4 5	a_1
IFHCDB (17,740 samples)	Step 2	2	1 3 4 5	a_2
CENPARMI (11,000 samples)	Step 3	3	1 2 4 5	a_3
	Step 4	4	1 2 3 5	a_4
	Step 5	5	1 2 3 4	a_5
The final result: $(a_1 + a_2 + a_3 + a_4 + a_5)/5$				

Table 2 Classification results using each model in HODA dataset

Classifier	Recognition rate (%)		
	Accuracy	Sensitivity	Specificity
VGG16	94.13	89.78	95.47
ResNet18	97.08	90.45	96.59
Xception	98.42	92.87	97.27
CBWME	99.87	96.52	99.20

It should also be mentioned that this idea is one of the advantages of the proposed idea because we are sure all the sets have been participated in test and training and that's why the obtained results are acceptable and significant. Finally, the final result can be obtained by averaging each result of each step, as shown in Table 1.

3.2 Classification results of CNN models and CBWME

The accuracy, sensitivity, and specificity results are provided in Tables 2, 3, and 4. From the result analysis in Table 2, it can be concluded that the Convolutional Bagging Weighted Majority Ensemble (CBWME) had the best results (99.87, 96.52, and 99.20%, respectively) during classification in HODA dataset, followed by the Xception model which obtained the accuracy, sensitivity, and specificity of 98.42, 92.87, and 97.27%, respectively. Then, the results obtained for accuracy, sensitivity, and specificity using ResNet18 are: 97.08, 90.45, and 96.59%. On the other side, VGG16 was the weakest model that attained 94.13, 89.78, and 95.47% accuracy, sensitivity, and specificity, respectively. When compared to ResNet18, the VGG16 specificity and sensitivity results don't differ very much, rather, they differ in performance accuracy by 97.08 and 94.13%, respectively.

In Table 3, the results related to IFHCDB are given, the results obtained for accuracy, sensitivity and specificity using VGG16 are 93.04, 87.49, and 92.20%, which has the weakest results. Instead, the results obtained for accuracy and specificity using CBWME are 98.42, 94.30, and 98.86%, which had the best results. Since Xception and CBWME have better architecture and efficient parameters, they obtained better results in comparison with VGG16 and ResNet18. It is noteworthy

Table 3 Classification results using each model in IFHCDB dataset

Classifier	Recognition rate (%)		
	Accuracy	Sensitivity	Specificity
VGG16	93.04	87.49	92.20
ResNet18	94.39	89.20	93.21
Xception	96.27	90.30	95.10
CBWME	98.42	94.30	98.86

Table 4 Classification results using each model in CENPARMI dataset

Classifier	Recognition rate (%)		
	Accuracy	Sensitivity	Specificity
VGG16	92.40	86.14	91.18
ResNet18	93.25	88.57	91.86
Xception	95.67	89.64	94.62
CBWME	97.13	90.87	97.29

that sensitivity obtained weaker results than accuracy and specificity. For example, the average results for accuracy, sensitivity, and specificity are 95.39, 89.64, and 94.41%, respectively.

The results of CENPARMI dataset are given in Table 4. According to Table 4, we can see that sensitivity and accuracy have obtained the weakest and strongest results. The average results obtained for accuracy, sensitivity, and specificity are 94.42, 88.46, and 93.45%, respectively. It should be mentioned that this dataset is older than mentioned ones and has less quality. The results obtained for accuracy, sensitivity, and specificity using VGG16 are 92.40, 86.14, and 91.18%, which were the weakest. In contrast, the results obtained for accuracy, sensitivity, and specificity using CBWME are 97.13, 90.87, and 97.29%, which had the best results. The results related to Xception are somehow close to CBWME. Especially about sensitivity which is 89.64% for Xception and 90.87% for CBWME. The values for accuracy in Xception and CBWME are 95.67 and 97.13%, respectively.

Farsi handwritten digits are very similar in writing to Arabic handwritten digits, but there are slight differences in the way the digits 0, 2, 4, 5, 6 are written. Despite the differences, we examined all types of existing writings in this study so the research of the mentioned project becomes valid in Arabic handwritten digits and provides significant results in case of testing with different datasets. Considering that in all datasets used in this research, including HODA, CENPARMI and IFHCDB, there are many examples of Arabic handwritten digits, but the focus of research is on Farsi handwritten digits.

Writing zero in Farsi is like English, which is a hollow circle but in Arabic zero is a solid circle. Writing one in each the two languages of Farsi and Arabic is the same and is almost the same as English. Writing two is different in Farsi and Arabic. In other words, writing 2 in Arabic is easier than writing it in Farsi because in Arabic 2 is generated from the intersection of a long vertical line and a short horizontal one.

While in Farsi, it is formed by a long vertical line and a parabolic curve with upward reasoning. In the case of number 3, the way of writing in Farsi and Arabic is the same.

In the case of number 4, the writing is completely different in both languages. In other words, the way of writing it in Farsi is much easier than in Arabic.

Writing 5 is different in both languages, except that the way it is written in Arabic is somewhat identical to a relatively large sloping hollow circle, which is somewhat similar to Farsi, but writing 5 in Farsi is relatively complex. However, it is much easier to identify.

There are two kinds of writing for number 6. The type of writing in Arabic is easier than in Farsi. The situation is completely different in numbers 7, 8 and 9, which means in both Farsi and Arabic the spelling is the same. Numbers 7 and 8 are composed of transversal lines, but with the difference that in the case of the number seven lines are moving from top to bottom, but in number 8 it is quite opposite, meaning that the lines are moving from bottom to top. The way of writing numbers 7 and 8 in Farsi is 7 and 8. Writing number 9 is very similar to the way of writing the number 9 in English, with the difference that in Farsi and Arabic, the last curvature related to the straight line isn't written. In other words, in Farsi, number nine consists of a circle and a relatively long vertical line to which it is attached.

Figure 3 shows the results obtained by the VGG16 on HODA, IFHCDB and CENPARMI datasets. According to the table presented in Fig. 3, we can see the best results are for numbers 1, 7 and 8 and the worst results are for 3. For example, the best result belongs to digit 7 from HODA dataset with amount of 91.45% and digit 3 from CENPARMI dataset has got the weakest result with amount of 86.09%.

Figure 4 shows the results by ResNet18 on HODA, IFHCDB and CENPARMI datasets. The best results are for 8 and the worst results are for 3. About HODA dataset, it should be mentioned that the weakest results were for number 2, which was 92.85%. The weakest results are for number 3 in CENPARMI dataset, which is very different with its following result related to zero (88.56%). The

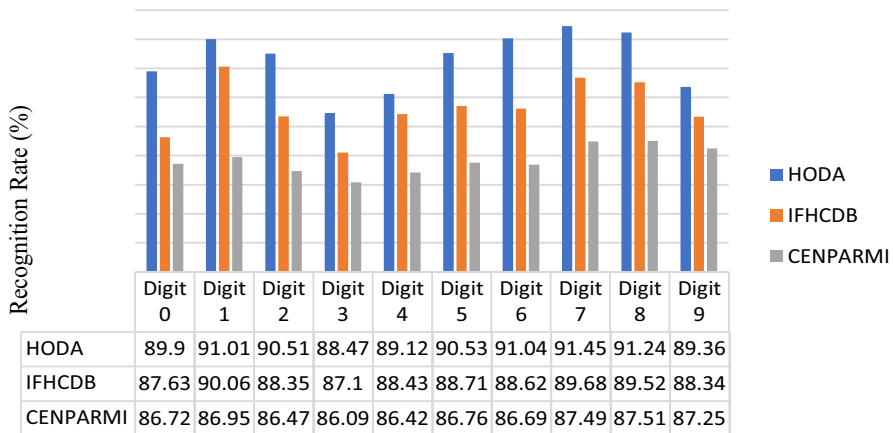


Fig. 3 Classification of digits using VGG16 model

same situation is also true about 3 in IFHCDB dataset. In other words, the results obtained for 3 are weaker than others, which is considered outlier data. If we put it aside from calculations, the averages will improve greatly. Number 5 had the best results, which was graded second in HODA dataset, and we can see this growing trend in other datasets for this number. In general, we can claim that ResNet18 had the best results for other digits except 2 and 3.

The results related to Xception method are given in Fig. 5. About HODA and IFHCDB datasets, digit of 8 has obtained the best results. However, digit of 7 has obtained the best results in dataset of CENPARMI. The weakest result is for number 3 in each of them. In a way that the weakest result is related to digit 3 in CENPARMI dataset using Xception method with amount of 88.13%. Considering the obtained results, we can say that even the results related to CENPARMI dataset have been better than previous methods. It should be mentioned that Xception is a new model in very deep convolutional neural networks and has shown acceptable results.

Here, the results obtained from the proposed method which is called Convolutional Bagging Weighted Majority Ensemble (CBWME) are analysed. You can see them in Fig. 6. The result obtained using CBWME is acceptable and better than other methods. In HODA and CENPARMI datasets, the best results are for 8; however, in CENPARMI, 7 gained the best results. The weakest results for all three datasets are for number 3. The interesting point is that the data related to 3 in all three datasets are considered outlier and vary greatly with other results. It is noteworthy saying that some digits like 0, 2, 4, 5 and 6, which had 2 spellings, showed good results and had moderate distance with the best results. The best thing about CBWME method is that there isn't great difference in the weakest and the best results, which indicates that the proposed method has a great ability to identify. Even datasets with average quality are analysed accurately, and the considered data are extracted and used where they are needed. The average values related to HODA, IFHCDB, and CENPARMI datasets are 97.65, 96.89, and

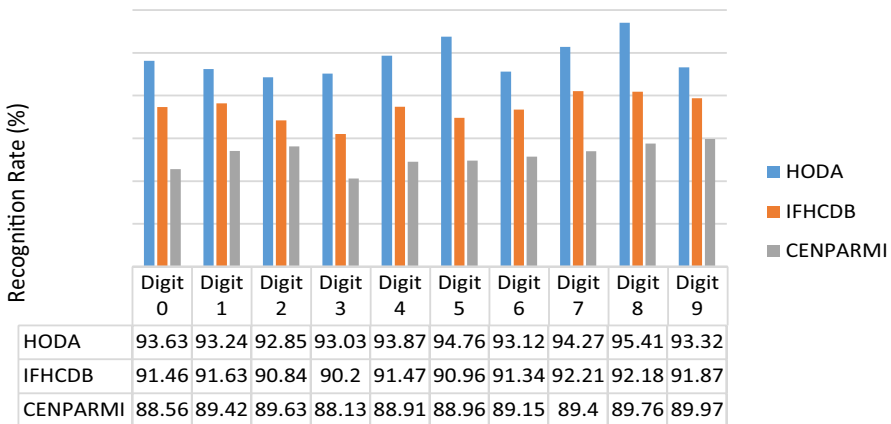


Fig. 4 Classification of digits using ResNet18 model

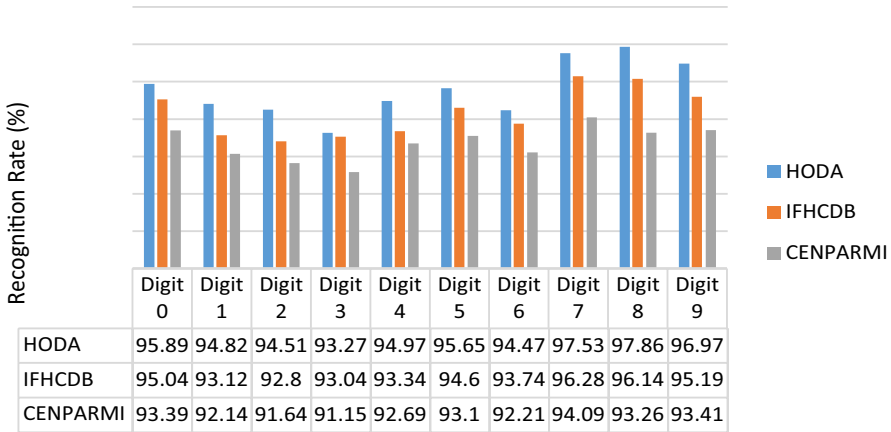


Fig. 5 Classification of digits using Xception model

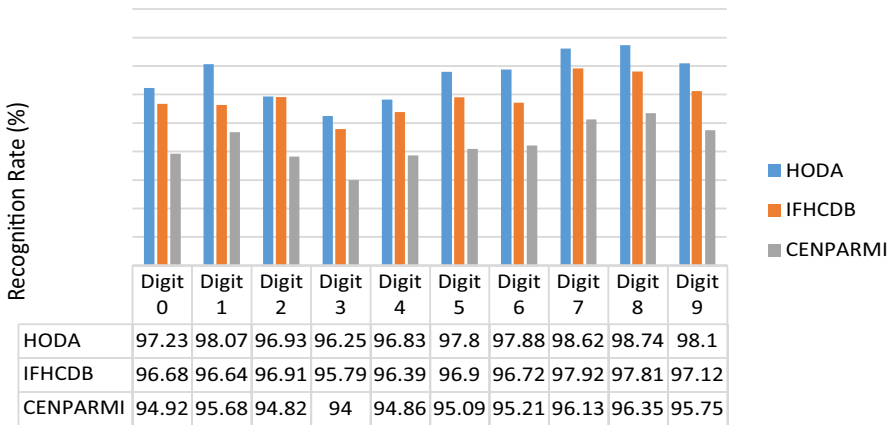


Fig. 6 Classification of digits using CBWME

95.28%, respectively (Fig. 7). In short, from the above analysis, it can be claimed that the CBWME method has the high ability in recognition of Farsi handwritten digits.

3.3 Comparison digit results of CNN models and CBWME

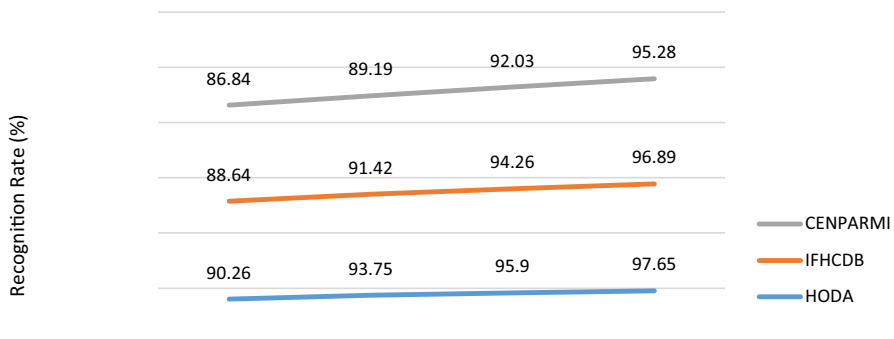
The average results of every dataset are shown in Fig. 7. From the result analysis, we can notice that the CBWME model had the best result in all of the dataset. The achieved results for datasets of HODA, IFHCDB, and CENPARI are, respectively: 97.65, 96.89, and 95.28% for average recognition rate. On the other hand, VGG16 model got the weakest results. Its results for HODA, IFHCDB, and CENPARI datasets are, respectively, 90.26, 88.64, and 86.84% for average recognition rate. The

obtained results for ResNet18 model are better than VGG16, but its results have a little more distance with Xception method. The obtained results of ResNet18 with different datasets are as follows: HODA: 93.75%, IFHCDB: 91.42%, and CENPARMI: 89.19% for average recognition rate. Xception model has got really good results, but it ranked the second place after the CE method. Xception model got 95.90, 94.26, and 92.03% with datasets of HODA, IFHCDB, and CENPARMI for average recognition rate, respectively. Finally, we can strongly claim that the CBWME has the best performance in all the used datasets.

Since the presented research can be evaluated precisely, we tried to present its running time in Fig. 8. The results in Fig. 8 indicate that the CBWME model has the longest running time. In other words, it has the worst results considering running time. It should be mentioned that there are a lot of differences between this model and the other proposed method, which comes after that. Simply, if the running time is a very important and effective component, CBWME is not a good model at all. Despite expectations, Xception has the second position higher than ResNet18 and the proposed method, which is an advantage for Xception. The best model is VGG16 considering running time.

3.4 Comparison results with other studies

In this section, we have attempted to bring some results about the newest scientific research in the field of Farsi handwritten digit recognition published in reliable scientific magazines and make contrast between them with the results from CBWME. It should be noted that traditional classifiers and CNNs have been used in most of the articles related to handwritten digit recognition or in other words, articles containing



	VGG16	ResNet18	Xception	CBWME
CENPARMI	86.84	89.19	92.03	95.28
IFHCDB	88.64	91.42	94.26	96.89
HODA	90.26	93.75	95.9	97.65

Fig. 7 The average results obtained for all the used datasets and all the models

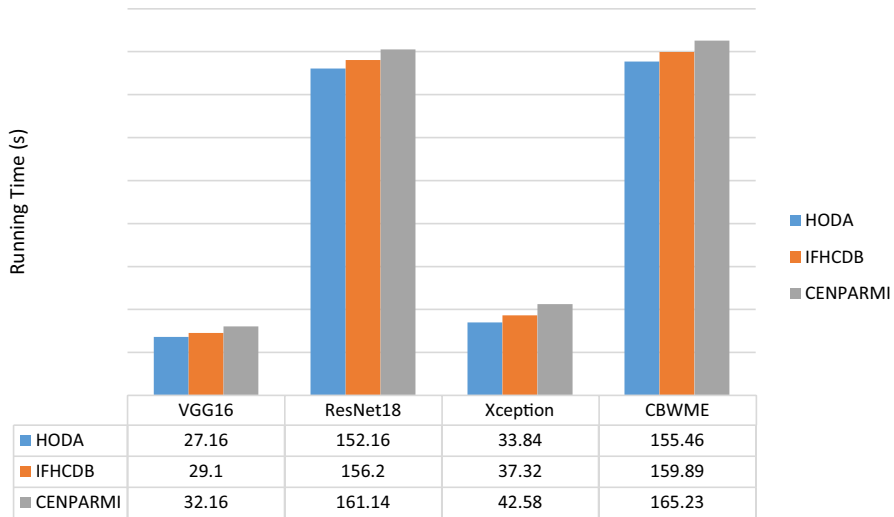


Fig. 8 The running time of the different models

VGG, ResNet, Inception, and Xception, and other new methods to recognize Farsi handwritten characters are not a lot (almost none).

It is better to remind that despite differences between Farsi and Arabic, they share some similarities in writing styles. Due to the lack of sources related to Farsi handwritten digit recognition using new methods and also similarities between Arabic and Farsi in writing styles, we tried to bring the results of some articles related to Arabic and compare the results with the proposed method to demonstrate the sensitivity of system better. The results related to the newest articles published in valid scientific magazines are brought here so that by comparing them with the proposed method, efficiency of the proposed method would become more observable. For example, some articles published in recent years are as follows (Table 5):

In this part, the results obtained from the proposed method are compared and contrasted with other works. Table 5 shows that the proposed result has better performance than others. It should be mentioned that the purpose of the study is to highlight the significance of ensemble learning methods using new techniques like DNN in order to get better results in Farsi handwritten digits recognition. To do such a research, we used three standard and commonly used datasets in Farsi, which are HODA, IFHCDB and CENPARMI. The best results were obtained from CBWME method using the HODA, IFHCDB and CENPARMI datasets: 99.87, 98.42, and 97.13%, respectively, as given in Table 5.

Then, Farahbakhsh et al. [10] made some changes in the architecture of CNN and achieved 99.67% using HODA dataset; Parseh et al. [18] presented a model of CNN and nonlinear multi-class and could achieve 99.56% by means of new changes in its architecture and structure; Modhej et al. [17] proposed a combined robust model for recognition of Farsi handwritten digits by which they achieved 99.55% using HODA dataset; Nanekaran et al. [12] proposed a model based on CNN based on which the

Table 5 Comparison of the proposed method and other researches

Research	Dataset	Technique	The highest accuracy (%)	The average accuracy (%)
Proposed method	HODA	CBWME (fivefold cross-validation)	99.87	99.34
		CBWME (separate test and train data)	99.07	98.71
	IFHCDB	CBWME (fivefold cross-validation)	98.42	97.69
		CBWME (separate test and train data)	97.40	96.81
	CENPARMI	CBWME (fivefold cross-validation)	97.13	96.36
		CBWME (separate test and train data)	96.27	95.92
Farahbakhsh [10]	HODA	DNN	99.67	–
Parseh [18]	HODA	CNN + Nonlinear SVM	99.56	–
Modhej [17]	HODA	Combination of different methods	99.55	–
Nanehkar et al. [12]	HODA	Directed CNN	99.45	–
Safarzadeh [16]	HODA	CNN + RNN	99.375	–
Akhlaghi [13]	HODA	CNN	99.34	–
Latif [11]	HODA	DNN	99.20	–
Elkhayati [14]	IFHCDB	CNN	97.40	–
Sahlol [15]	CENPARMI	Combination of different methods	96.00	–

result was 99.45% using HODA dataset; Safarzadeh and Jafarzadeh [16] proposed a combined method based on CNN and RNN and could achieve 99.375% considering the innovations in new method; Akhlaghi and Ghods [13] proposed a method for Farsi handwritten digits recognition based on which the main idea of CNNs developed. Authors have tried to recognize string digits like handwritten phone numbers and used different methods to achieve this goal. The best result using HODA dataset was 99.34%; Latif et al. [11] presented a model based on DNN and got 99.20%; Elkhayati et al. [14] have used directed CNN to recognize handwritten digits, and IFHCDB is used which resulted in 97.40%; Sahlol et al. [15] presented a robust mixed model using CENPARMI. The method was robust and had high proficiency. The authors could achieve 96.00% using this method. In Table 5, the results of some papers related to the recognition of Farsi and Arabic digits are shown. Majority of the presented results are almost the same with the proposed CBWME method, which reveal its efficiency. Despite the efficiency and state-of-the-art methods (when they were introduced) of other methods presented in Table 5, they are not as efficient as the proposed results. In other words, they are not comparable with CBWME.

4 Discussion

From the above analysis and comparison, CBWME has gained the best results, following Xception, ResNet18, and VGG16 model in recognition of both handwritten Farsi characters and digits, which were caused by the following reasons:

First of all, using an architecture with very small (3×3) convolution filters, VGG16 enhances a thorough evaluation of networks of increasing depth. This is

an indication that there is a notable improvement on the prior-art configurations that can be achieved by pushing the depth to 16 to 19 weight layers. It can express the dataset characteristics more accurately during image identification and classification. However, as the network goes deep, the number of model parameters and the complexity of calculations during training increase, which, in turn, results in more training time and low training efficiency.

With ResNet18, up to thousands of residual layers can be utilized in creating a network and then trained. This is different from normal sequential networks, where as you increase the number of layers, there are less improvements in the network performance. A few more novel techniques which ResNet18 introduced are: (1) use of standard stochastic gradient descent (SGD) along with a reasonable initialization function which keeps the training intact. (2) Changes in pre-processing the input, where the input is first divided into patches and then fed into the network.

The Xception module is robust, very strong, and can process correlations of cross-channels and spatial relations with maps fully decoupled. The Xception architecture is a linear stack of depth-wise separable convolution layers with residual connections. This makes the architecture very easy to define and modify; it takes only 30–40 lines of code using a high-level library such as Keras or TensorFlow-Slim, not unlike an architecture such as very deep convolutional neural networks.

By using the central limit theorem (which involves averaging of many classifiers), the bagging strategy can overcome or avoid over-fitting successfully. The randomization process enhances the robustness of the models. Thus, bagging is aimed at improving classification by combining single classification, and the results are better than one single classification. CBWME model is adopted the bagging strategy and has its advantages and thus reaches the best results compared with VGG16, ResNet18, and Xception.

5 The hypothesis and limitations of CBWME

CBWME is a powerful ensemble method, which helps to reduce variance and, by extension, prevent over-fitting. It improves model precision by using a group (or "ensemble") of models which, when combined, outperform individual models when used separately.

However, its limitations are as follows: 1. It is giving its final prediction based on the mean predictions from the subset trees, rather than outputting the precise values for the classification or regression model. 2. It introduces a loss of interpretability of a model. The resultant model can experience lots of bias when the proper procedure is ignored. Despite CBWME being highly accurate, it can be computationally expensive, and this may discourage its use in certain instances. 3. It may result in high bias if it is not modelled properly and thus may result in under-fitting. 4. Since we must use multiple models, it becomes computationally expensive and may not be suitable in various use cases.

6 Conclusion

In order to improve the recognition accuracy of Farsi handwritten digit recognition, in this work, the novel CBWME network structure model based on convolution bagging weighted majority ensemble learning was developed by integrating the convolution neural network (CNN) and bagging weighted majority ensemble learning. In this model, the VGG16, ResNet18, and Xception CNN models were designed as base classifiers, while the bagging weighted majority ensemble learning was explored in combining the base classifiers results, which were later used in identifying Farsi handwritten digits. The performance of the CBWME model was evaluated based on three databases: HODA, IFHCDB and CENPARMI, and its results were compared to those of the CNN models. From the experimental result analysis, it was observed that the proposed CBWME model achieved the best average recognition accuracy (97.65%), followed by the Xception model (95.9%), ResNet18 model (93.75%), and VGG16 model (90.26%) in HODA dataset. The accuracy orders were the same as in IFHCDB and CENPARMI datasets. Moreover, in handwritten letter recognition, CBWME obtained higher recognition accuracy than the other three CNN models. Thus, to this end, the proposed model brings new insight to the field and can be a very useful tool in Farsi handwritten digits recognition. However, it costs amount computational time to obtain the higher accuracy. In the future, many applications of the number recognition and handwriting character double the importance of this topic.

7 Recommendation

The research conducted in the field of Farsi handwriting digits recognition so far is very limited. There are good outcomes in spite of little researches. But there is still a need for more work in this area. Many applications of the number recognition and handwriting character double the importance of this topic [37, 38]. Future work recommends the use of new features according to the figure and content of the image, the use of image derivatives, and the extraction of features that are resistant to image rotations in different directions. It also classifies images using various online and offline data using clustering and other classification methods. Combined classifiers as well as changes in deep neural network parameters and other classifiers and finding the optimal value of parameters for them can be helpful in this regard [39, 40]. Providing methods for detecting digits and characters in noisy environments can also be an interesting topic. Recognition of Farsi handwriting, especially in handwritten strings, still requires much research. Considering the long length and diversity of subject, we tried to combine the mentioned idea with new methods and techniques and test it. For example, the present idea was used to test longer digits like phone numbers, ZIP codes, ID numbers and student ID numbers and demonstrate the output as a text. This idea was used for both letters and digits. Considering the subject and recent advancements in

the field of artificial intelligence especially in deep neural networks, providing mixed and innovative ideas is common.

Acknowledgements The writers want to appreciate Pzcnnet Ltd. (<http://www.pzcnnet.com/>). The author also likes to appreciate all the judges and editors whose useful suggestions helped improve the article.

Funding This work is partly supported by grants from the National Natural Science Foundation of China (Project no. 61672439) and the Fundamental Research Funds for the Central Universities (#20720181004).

References

1. Sajedi H, Bahador M (2016) Persian handwritten number recognition using adapted framing feature and support vector machines. *Int J Comput Intell Appl* 15(1):165–180
2. Karimi H, Esfahanimehr A, Mosleh M, Salehpour S, Medhati O (2015) Persian handwritten digit recognition using ensemble classifiers. *Procedia Comput Sci* 73(1):416–425
3. Khosravi H, Kabir E (2006) Introduction of Two Fast and Effective Features for Handwritten Farsi Digit Recognition. In: 2006 4th Iranian Conference on Machine Vision and Image Processing (MVIP), pp 87–95
4. Jain V, Dubey A, Gupta A, Sharma S (2016) Comparative Analysis of Machine Learning Algorithms in OCR. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, pp 1089–1092
5. Chaudhary D, Sharma K (2019) Hindi Handwritten Character Recognition Using Deep Convolution Neural Network. In: 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, pp 961–965
6. Al-Janabi S, Alkaim A (2020) A nifty collaborative analysis to predicting a novel tool (DRFLLS) for missing values estimation. *Soft Comput* 24(1):555–569
7. Kilvisharam OMA, Poruran S, Caffiyar MY (2020) A novel deep convolutional neural network architecture based on transfer learning for handwritten Urdu character recognition. *VJESN* 27(4):1160–1165
8. Granet A, Morin E, Mouchère H, Quiniou S, Viard-Gaudin C (2018) Transfer Learning for Handwriting Recognition on Historical Documents. In: 2018 International Conference on Pattern Recognition Application & Methods (ICPRAM), pp 432–439
9. Jaderberg M, Simonyan K, Vedaldi A, Zisserman A (2014) Deep structured output learning for unconstrained text recognition. *arXiv preprint <http://arXiv:1412.5903>*
10. Farabakhsh E, Kozegar E, Soryani M (2017) Improving Persian Digit Recognition by Combining Data Augmentation and AlexNet. In: 2017 10th Iranian Conference on Machine Vision and Image Processing (ICMVIP), pp 265–270
11. Latif G, Alghazo J, Alzubaidi L, Naseer MM, Alghazo Y (2018) Deep Convolutional Neural Network for Recognition of Unified Multi-Language Handwritten Numerals. In: 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR). IEEE, pp 90–95
12. Nanekaran YA, Zhang D, Salimi S, Chen J, Tian Y, Al-Nabhan N (2020) Analysis and comparison of machine learning classifiers and deep neural networks techniques for recognition of farsi handwritten digits. *J Supercomput* 1(3):1–30
13. Akhlaghi M, Ghods V (2020) Farsi handwritten phone number recognition using deep learning. *SN Appl Sci* 2(3):1–10
14. Elkhayati M, Elkettani Y (2020) Towards directing convolutional neural networks using computational geometry algorithms: application to handwritten Arabic character recognition. *Adv Sci Technol Eng* 5(5):137–147
15. Sahlol AT, Elaziz M, Qaness MA, Kim S (2020) Handwritten Arabic optical character recognition approach based on hybrid whale optimization algorithm with neighborhood rough set. *IEEE Access* 8(2):23011–23021
16. Safarzadeh VM, Jafarzadeh P (2020) Offline Persian Handwriting Recognition with CNN and RNN-CTC. In: 25th International Computer Conference Computer Society of Iran (CSI), pp 1–10
17. Modhej N, Bastanfard A, Teshnehlab M, Raiesdana S (2020) Pattern separation network based on the hippocampus activity for handwritten recognition. *IEEE Access* 8:212803–212817

18. Parseh M, Rahmanimanesh M, Keshavarzi P (2020) Persian handwritten digit recognition using combination of convolutional neural network and support vector machine methods. *Int Arab J Inf Technol* 17(4):572–578
19. Janabi S, Mahdi MA (2019) Evaluation prediction techniques to achievement an optimal biomedical analysis. *Int J Grid Util Comput* 10(5):512–527
20. Moradi V, Razzazi F, Behrad A (2016) Recognition of handwritten persian two-digit numerals using a novel hybrid SVM/HMM algorithm. *Majlesi J Electr Eng* 10(3):233–245
21. Parvin H, Alizadeh H, Bidgoli B (2010) A new divide and conquer based classification for OCR. *Converg Hybrid Inform Technol* 3(1):1–1
22. Janabi S, Alkaim AF, Adel Z (2020) An innovative synthesis of deep learning techniques (DCapsNet & DCOM) for generation electrical renewable energy from wind energy. *Soft Comput* 24(14):10943–10962
23. Oh B, Lee J (2018) A Case Study on Scene Recognition Using an Ensemble Convolution Neural Network. In: 2018 20th International Conference on Advanced Communication Technology (ICACT), pp 351–353
24. Agahi H, Mahmoodzadeh A, Salehi M (2018) Handwritten digits recognition using an ensemble technique based on the firefly algorithm. *J Inf Syst Telecommun* 3(1):123–136
25. Ameryan M, Schomaker L (2019) A limited-size ensemble of homogeneous CNN/LSTMs for high-performance word classification. *arXiv preprint arXiv: 1912.03223*
26. Janabi S, Mohammad M, Sultan A (2020) A new method for prediction of air pollution based on intelligent computation. *Soft Comput* 24(1):661–680
27. Janabi S, Shourbaji I, Shojafar M, Abdelhag M (2017) Mobile Cloud Computing: Challenges and Future Research Directions. In: 2017 10th International Conference on Developments in E-Systems Engineering, IEEE, pp 62–67
28. Mahdi MA, Janabi S (2019) A Novel Software to Improve Healthcare Base on Predictive Analytics and Mobile Services for Cloud Data Centers. In: International Conference on Big Data and Networks Technologies, pp 320–339
29. Goswami S, Passi K (2019) Real time Static Gesture Detection Using Deep Learning. In: 2019 7th International Conference on Big Data Analytics, pp 408–426
30. Hoffman J (2019) Cramnet: layer-wise deep neural network compression with knowledge transfer from a teacher network. *arXiv preprint arXiv: 1904.05982*
31. He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp 770–778
32. Liu Y, Yang J (2019) Face Recognition Method Based on Convolutional Neural Network. In: International Conference in Communications, Signal Processing, and Systems (CSS), pp 1925–1929
33. Chollet F (2016) Xception: Deep Learning with Depth-wise Separable Convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1251–1258
34. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning Deep Features for Discriminative Localization. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2921–2929
35. Jinsakul N, Tsai CF, Tsai CE, Wu P (2019) Enhancement of deep learning in image classification performance using xception with the swish activation function for colorectal polyp preliminary screening. *Mathematics* 7(12):1170–1182
36. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) MobileNets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv: 1704.04861*
37. Dehghanian A, Ghods V (2018) Farsi Handwriting Digit Recognition Based on Convolutional Neural Networks. In: 2018 6th International Symposium on Computational and Business Intelligence (CBI), IEEE, pp 65–68
38. Alizadehashraf B, Roohi S (2017) Persian Handwritten Character Recognition Using Convolutional Neural Network. In: 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP), pp 247–251
39. Rani NS, Subramani AC, Kumar A, Pushpa BR (2020) Deep Learning Network Architecture Based Kannada Handwritten Character Recognition. In: 2020 Second International Conference on Inventive Research in Computing Applications (IRCA), IEEE, pp 213–220
40. Le CY, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444

Authors and Affiliations

Y. A. Nanehkaran^{1,2} · Junde Chen¹ · Soheil Salimi³ · Defu Zhang¹

Y. A. Nanehkaran
artavil20@gmail.com

Soheil Salimi
soheil.salimi.n@gmail.com

¹ School of Informatics, Xiamen University, Xiamen 361005, Fujian, People's Republic of China

² School of Information Engineering, Jiangxi University of Science and Technology, 86, Hongqi Ave, Ganzhou, Jiangxi, People's Republic of China

³ Department of Electrical and Computer Engineering, Seraj Institute of Higher Education, Tabriz, Iran