Image feature extraction for handwritten digits recognition using LeNet

MD. Samiul Hauque Chowdhury (18-36072-1), MD. Sabbir Hasan (17-33918-1), Mehedi Hasan Joy (18-38413-2),Shihab Hossain(18-38131-2),

Fahad, Abdullah Al (18-37904-2)

*Department of Computer Sciences, American International University-Bangladesh*

**Abstract**

Handwritten digits recognition has been treated as a multi-class classification problem in the machine learning context, where each of the ten digits (0–9) is viewed as a class and the machine learning task is essentially to train a classifier that can effectively discriminate the ten classes. In practice, it is very usual that the performance of a single classifier trained using a standard learning algorithm is varied on different datasets, which indicates that the same learning algorithm may train strong classifiers on some datasets but weak classifiers may be trained on other datasets. It is also possible that the same classifier shows different performance on different test sets, especially when considering the case that image instances can be highly diverse due to the different handwriting styles of different people on the same digits. To address the above issue, development of ensemble learning approaches have been very necessary to improve the overall performance and make the performance more stable on different datasets. In this paper, we propose a framework that involves CNN-based feature extraction from the MINST dataset and algebraic fusion of multiple classifiers trained on different feature sets, which are prepared through feature selection applied to the original feature set extracted using CNN. The experimental results show that the classifiers fusion can achieve the classification accuracy of ≥ 98%.

**Keywords** Machine learning · Ensemble learning · Classification · Classifiers fusion · Random forests · Granular computing

# Introduction

Handwriting digits recognition refers to the process of transforming the ordered trajectory generated by writing on handwriting equipment into the internal code of digits. It is actually a mapping process from the coordinate sequence of handwritten trajectory to the internal code of digits. It is one of the most natural and convenient means of human–computer interaction. With the popularity of mobile information tools such as smartphones and handheld computers, hand- written digits recognition technology has entered the era of large-scale application. Handwritten digits recognition enables users to input text in the most natural and convenient way. It is easy to learn and use, and can replace keyboards or mouses. There are many kinds of devices for handwriting inputs, such as electromagnetic induction handwriting boards, pressure-sensitive hand-writing boards, touch screens, touch panels, ultrasonic pens, etc. Handwriting dig- its recognition belongs to the category of digits recognition and pattern recognition. In terms of the recognition process, digits recognition can be divided into two categories: off-line recognition and on-line recognition. In terms of recognition objects, it can also be divided into two categories: handwriting digits recognition and print digits recognition.

Also, it is well known that the handwritten digits recognition is a challenging problem. In recent years, there are many algorithms proposed for handwritten digits recognition. Boukharouba ([2017](#_bookmark12)) develops a new feature extraction technique for handwritten digit recognition based on support vector machines (SVM). During this method, the vertical and horizontal directions of a digit image are combined with the famous Freeman chain code, and the approach does not require any normalization of digits.

In the machine learning context, it is commonly known that each standard learning algorithm usually shows different performance on different datasets. In other words, the use of an algorithm may lead to the production of strong classifiers on some datasets but the classifiers trained on other datasets using the same algorithm may be much weaker. In the case of handwritten digits recognition, a standard learning algorithm may be capable of learning some but not all specific characteristics of handwritten digits.

Also, the same classifier may show different performance on different datasets, due to the different data distribution. In addition, instances of handwritten digits usually show very diverse characteristics due to different handwriting styles of different people, even if the instances belong to the same class. The main contributions of this paper include: (1) the use of CNN to extract more diverse features from each hand- written digit image and different feature sets are prepared through filter-based feature selection; (2) an ensemble learning framework is proposed, which involves multi-level fusion of multiple classifiers trained on different feature sets using different learning algorithms.

# Related work

This section provides a review of the applications of convolutional neural networks for image classification, an overview of handwritten digits recognition and a review of traditional machine learning methods alongside potential improvements through the use of granular computing concepts.

Various methods have been proposed and high recognition rates are reported for the recognition of English handwritten digits [13–15]. Niu and Suen [13] proposed recognize handwritten digits using Convolutional Neural Network (CNN) and Support Vector Machine (SVM). There Experiments have been conducted on MNIST digit database. They achieve recognition rate of 94.40 % with 5.60 % with rejection. Tissera and McDonnell [14] introduced a supervised auto-encoder architecture based on extreme machine learning to classify Latin handwritten digits based on MNIST dataset. The proposed technique can correctly classify up to 99.19 %. Ali and Ghani introduced Discrete Cosine Transform based on Hidden Markov models (HMM) to classify handwritten digits. They used MNIST as training and testing datasets. HMM have been applied as classifier to classify handwritten digits dataset. The algorithm provides promising recognition results on average 97.2 %.

In recent years many researchers addressed the recognition of text including Arabic. In 2011, Melhaoui et al. [16] proposed an improved method for recognizing Arabic digits based on Loci characteristic. Their work is based on handwritten and printed numeral recognition. The recognition is carried out with multi-layer perceptron technique and K-nearest neighbor. They trained there algorithm on dataset contain 600 Arabic digits with 200 testing images and 400 training images. They were able to achieve 99 % recognition rate on small database. 568 A. El-Sawy et al. In 2008, Mahmoud [17] proposed a technique for the automatic recognition of Arabic handwritten digits using Gabor-based features and Support Vector Machines (SVMs). They used a medium database have 21120 samples written by 44 writers. The dataset contain 30 % for testing and the remaining 70 % of the data is used for training. They achieved average recognition rates are 99.85 % and 97.94 % using 3 scales & 5 orientations and using 4 scales & 6 orientations, respectively. In 2014, Takruri et al. [18] presented three level classifier based on Support Vector Machine, Fuzzy C Means and Unique Pixels for the classification of handwritten Arabic digits. They tested the new algorithm on a public dataset. The dataset contain 3510 images with 40 % are used for testing and 60 % of images are used for training. The overall testing accuracy reported is 88 %. In 2013, Pandi Selvi and Meyyappan [1] presented a method to recognize Arabic digits using back propagation neural network. The final result shows that the proposed method provides an recognition accuracy of more than 96 % for a small sample handwritten database.

# Proposed framework

In this section, we provide a description of CNN based feature extraction and present an ensemble learning framework that involves multi-level fusion of multiple classifiers trained on different feature sets using different learning algorithms. We also justify how the design of the proposed framework involves the application of granular computing concepts.

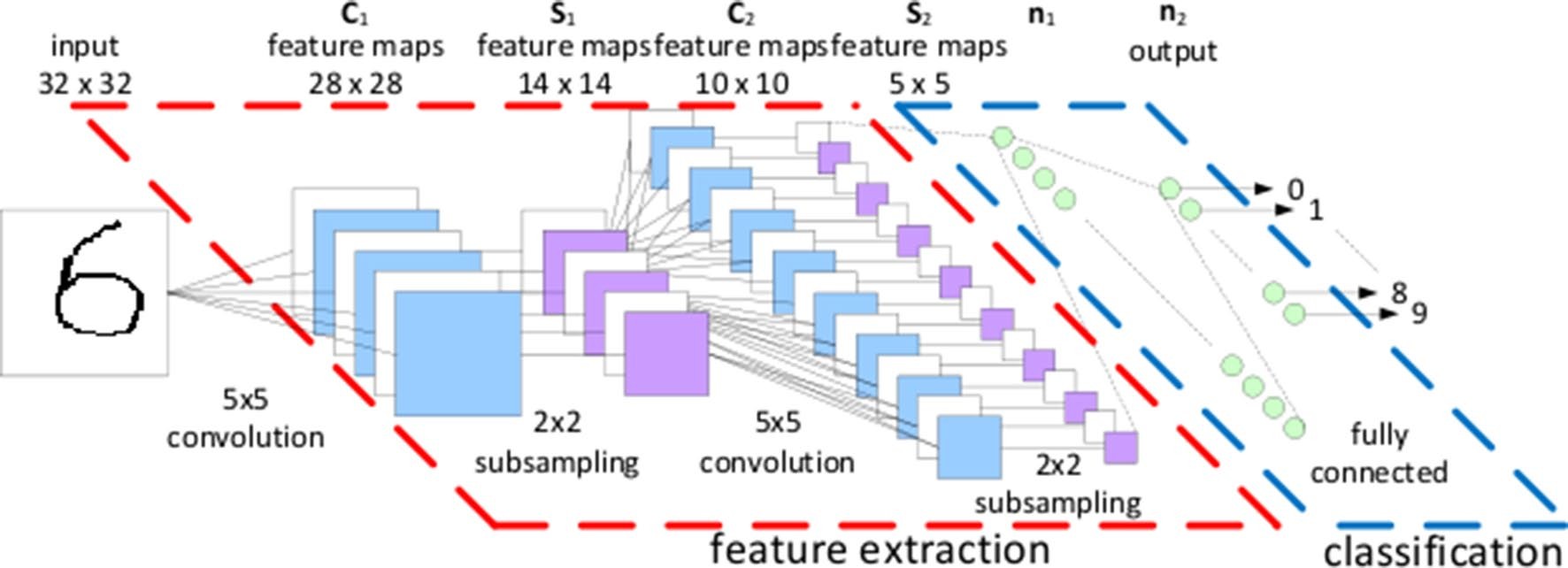
## **CNN feature -extraction**

During our method, we use CNNLeNet-5 to obtain more diverse features from each handwritten digit image. The proposed CNN feature extraction for handwritten digit images by LeNet-5 is illustrated in Fig. [1](#_bookmark2).

The LeNet architecture is considered as the first architec- ture for convolutional neural networks. We can easily see from the LeNet-5 in Fig. [1](#_bookmark2) that many feature maps are gen- erated in each layer. So, we can obtain more diverse features than using other common methods.

The LeNet-5 is an excellent architecture for handwritten digit recognition. The LeNet-5 has two parts, one is feature extraction, whereas the other one is classification which is used to classify objects. During our approach, we do not use the LeNet-5 to do classification (blue part in Fig. [1](#_bookmark2)), and we only use it to extract features from images. During the classification, we use the proposed ensemble learning framework instead of a neural network that consists of fully connected layers.

Given an image of 32 × 32 × 1, firstly, a convolution layer with six 5 × 5 filters with the stride of 1 is used and an out- put matrix of 28 × 28 × 6 is generated. With the stride of 1 and no padding, the feature map is reduced from 32 × 32 to 28 × 28. Then average pooling with the filter width of 2 and the stride of 2 is taken and the dimension is reduced by the factor of 2 and ends up with 14 × 14 × 6. Further- more, another convolution layer with sixteen 5 × 5 filters is used leading to an output matrix of 10 × 10 × 16. Then another pooling layer is involved and ends up with an out- put matrix of 5 × 5 × 16. Therefore, we extract sixteen 5 × 5 feature maps from each image, and each feature map (5 × 5)

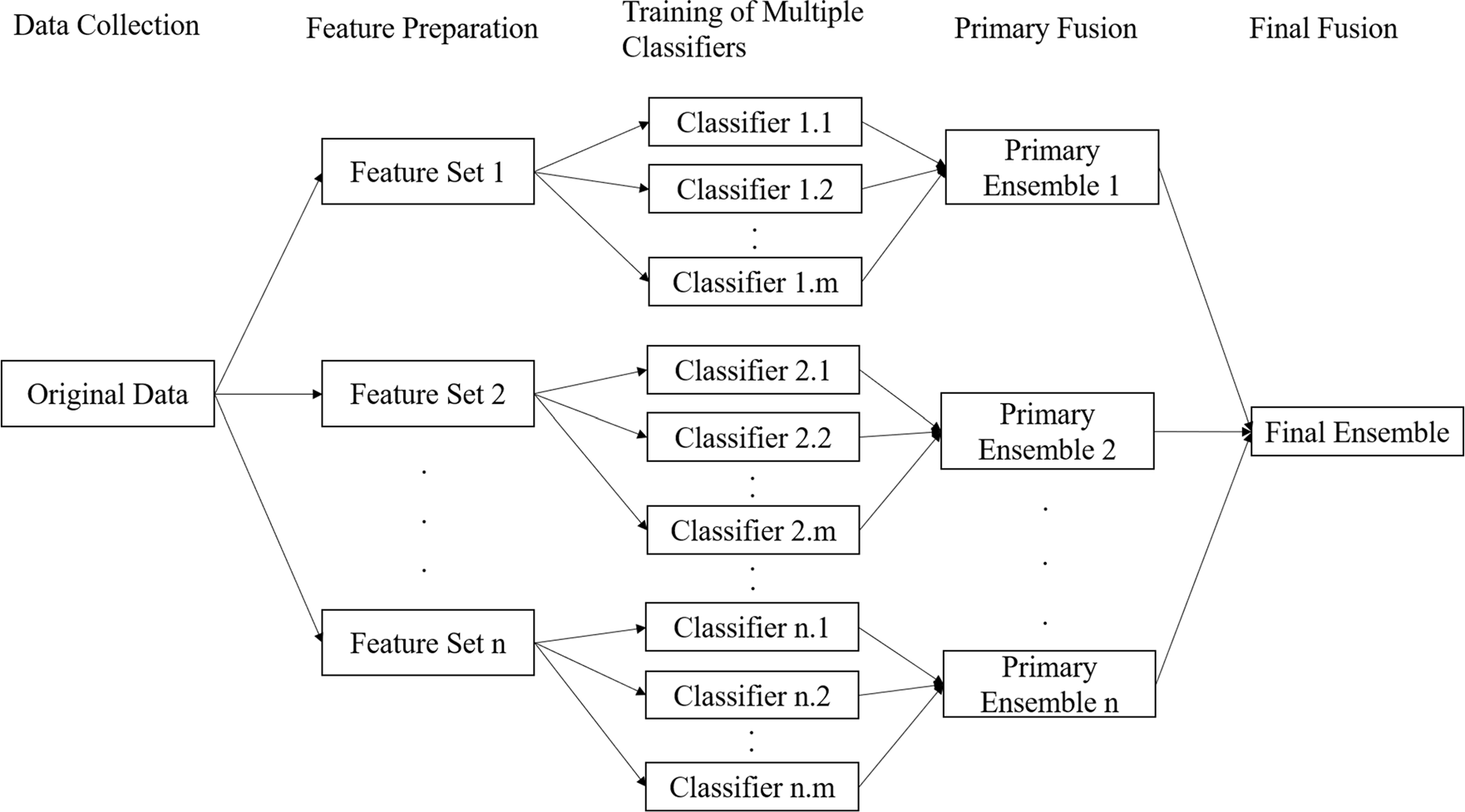


is treated as a column vector (25 × 1). Overall, there are two convolution layers, two subsampling layers, and two fully connected layers in the LeNet-5.

## **Multi**‑**level** **fusion** **of** **classifiers**

The proposed ensemble learning framework involves multi- ple levels of fusion of diverse classifiers trained on different feature sets. The entire procedure of the proposed framework is illustrated in Fig. [2](#_bookmark3).

In particular, as shown in the feature preparation layer in Fig. [2](#_bookmark3), different feature sets can be prepared through feature extraction using different methods (but we only obtain one feature set extracted using CNN in this paper). Also, the feature set can be combined to make up a primary ensemble, which is combined further with the other base classifiers to make up a secondary ensemble. In this context, a secondary ensemble can be created on each feature set and some or all of the secondary ensembles can be fused further to make up a higher level ensemble or even the final ensemble.



The proposed ensemble learning framework is essentially designed in the setting of granular computing, which is a formalized paradigm of information processing (Pedrycz [2011](#_bookmark44); Pedrycz and Chen [2011](#_bookmark45), [2015a](#_bookmark46), [b](#_bookmark47)). In general, granular computing can be considered as a method of structural thinking at the philosophical level but can also be used as a strategy of structural problem solving at the practical level (Yao [2005b](#_bookmark62)).

In theory, two main concepts of granular computing are referred to as granule and granularity (Liu and Cocea [2017](#_bookmark35), [2018](#_bookmark36); Liu et al. [2018](#_bookmark39)). Granule is defined as a collection of smaller particles that can form a larger unit. In the context of ensemble learning, each ensemble can be viewed as a granule since it consists of multiple classifiers. While gran- ules can be of very different sizes, the concept of granularity becomes highly needed to deal with the different sizes of different granules, that is, to involve different granules in different levels of granularity, according to the scale of their actual sizes. The proposed ensemble learning framework involves multiple levels of classifiers fusion, where each of the levels can be viewed as a specific level of granularity. In this context, a primary ensemble that only consists of base classifiers is viewed as a granule at the basic (bottom) level of granularity, whereas the final ensemble that may involve both base classifiers and lower level ensembles is viewed as a granule at the top level of granularity.

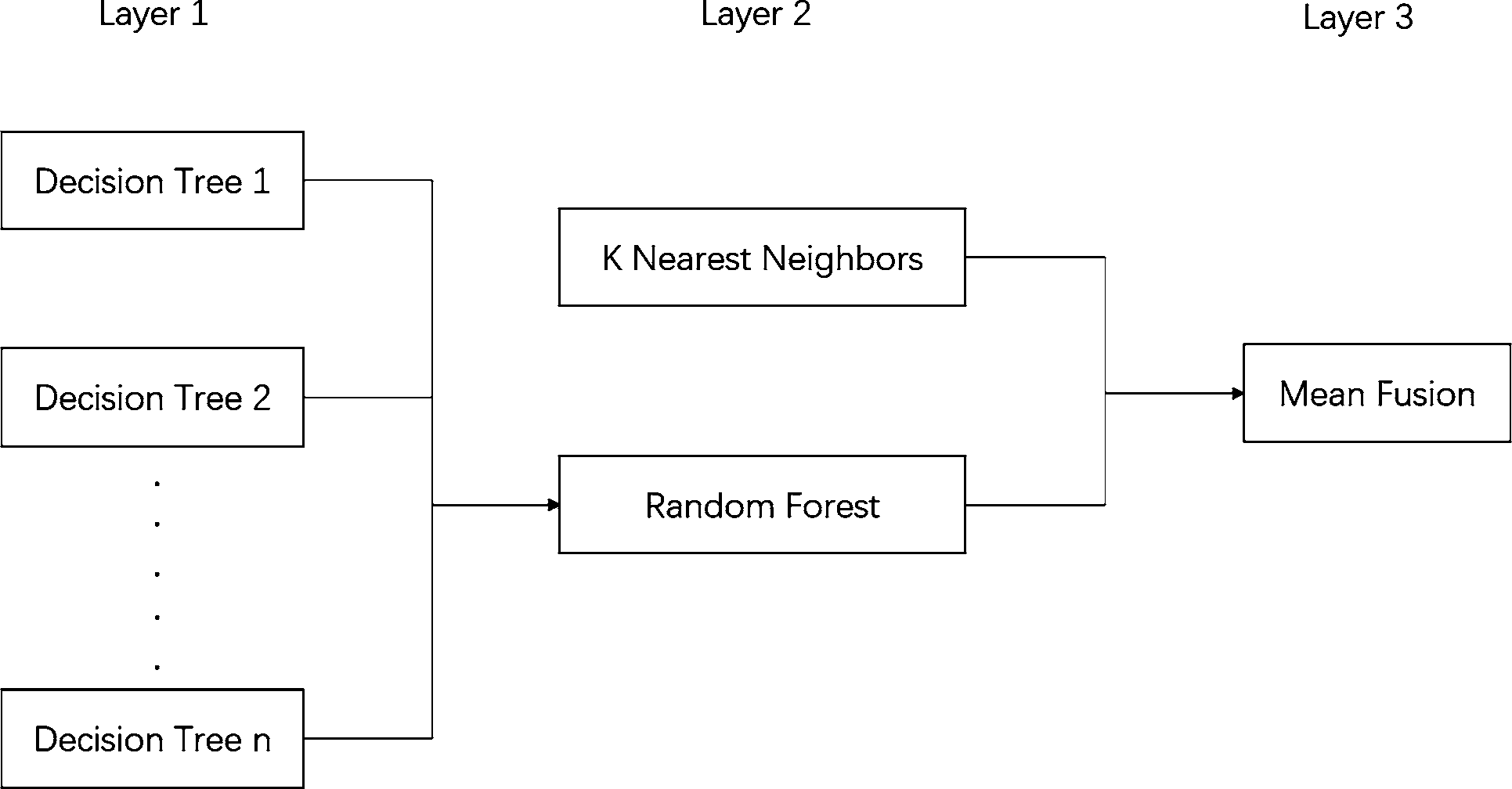
# Experimental results and discussion

In this section, we report an experimental study conducted on the MNIST dataset, which is essentially a 10-class (0–9) classification task in the setting of machine learning.

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 (Niu and Suen [2012](#_bookmark43)). The database consists of a training set of 60,000 images and a test set of 10,000 images.

In this experimental study, the whole procedure involves feature extraction, feature selection and training and fusion of classifiers. During the CNN feature extraction, we input each digit image to the LeNet-5, and output its feature maps in the 3rd layer (16 × 10 × 10). Furthermore, those feature maps are changed into a single column. In terms of the set- ting of the CNN architecture, the activation function is set as sigmoid, and the loss function is set as a mean squared error, and the optimization function is set as l2 regularizer. In addition, the input Batch size is set as 1. There are 60,000 images in the MNIST and the Epoch is 1, therefore, there are 60,000 iterations in total.

In the feature selection stage, we apply the correlation- based feature subset selection method (Hall and Smith [1997](#_bookmark31)) to obtain a reduced set of features. In this way, we use the reduced set of selected features alongside the original feature set extracted using CNN, such that diversity can be created through training classifiers on the two different feature sets. In the classifiers training and fusion stage, we adopt KNN and RF for training base classifiers and primary ensembles, respectively, on the two feature sets. On each feature set, a secondary ensemble is obtained through combining the base classifier (trained using KNN) and the primary ensemble of decision trees (created using RF). The two primary ensembles created on the two feature sets are combined further to make up a larger ensemble for final fusion. The whole setting of the ensemble creation on each feature set is illustrated in Fig. [3](#_bookmark5).



The results on the MINST dataset is shown in Table [1](#_bookmark6) in terms of classification accuracy. The results indicate that the nature of the KNN method through instance-based learning leads to the accuracy of ≥ 95.8% on the two feature sets. Also, the RF method is generally very capable of training highly diverse decision tree classifiers on different training samples and feature subsets, which leads to the accuracy of ≥ 95.7% on the two feature sets.

On the above basis, the further fusion of the base classifier (trained using KNN) and the decision tree (primary) ensemble (created using RF) leads to an improvement of the classification performance on each feature set, which indicates that the different learning strategies between the CNN and RF methods can really result in diversity between their trained classifiers. The final fusion of the above two secondary ensembles created on the two feature sets leads to a further improvement of the classification performance. In addition, although feature selection may not necessarily lead to advances in the classification performance for each single classifier trained on the reduced feature set in comparison with using the full feature set, the fusion of classifiers trained on the two feature sets can lead to an improve, which would indicate that the preparation of different feature sets through feature selection can effectively lead to the creation of diversity among classifiers trained on the

different feature sets.

Overall, the experimental results suggest that multilevel fusion of classifiers through various ways of diversity creation is encouraged towards advances in the classification performance in a layer-by-layer manner.

# Conclusions

In this paper, we have proposed a framework that involves CNN based feature extraction and multi-level fusion of diverse classifiers. In particular, we have designed to increase the diversity among classifiers through preparing different feature sets and using different learning algorithms for classifiers training. The experimental results show that our proposed ensemble approach can achieve the organization accuracy of ≥ 98% using the MNIST dataset and the results also indicate that the setting of ensemble learning which aims to train diverse classifiers is very useful to advance the overall performance of classification.

In future, we will investigate how to achieve optimal feature subsets selection to boost the performance further through using some optimization techniques.

# References

Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M (2017a) Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network. Inf Sci 405:81–90

Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H (2017b) Deep con- volutional neural network for the automated detection and diagno- sis of seizure using eeg signals. Comput Biol Med 100:270–278

Amor NB, Benferhat S, Elouedi Z (2004) Naive bayes vs decision trees in intrusion detection systems. In: ACM symposium on applied computing, pp 420–424

Biau G, Devroye L, Lugosi G (2009) Consistency of random forests and other averaging classifiers. J Mach Learn Res 9(1):2015–2033

Breiman L (2001) Random forests. Mach Learn 45(1):5–32 Chatterjee K, Kar S (2017) Unified granular-number-based AHP-

VIKOR multi-criteria decision framework. Granul Comput 2(3):199–221

Chen SM, Chang YC (2011) Weighted fuzzy rule interpolation based on ga-based weight-learning techniques. IEEE Trans Fuzzy Syst 19(4):729–744

Chen SJ, Chen SM (2001) A new method to measure the similarity between fuzzy numbers. In: IEEE international conference on fuzzy systems, Melbourne, pp 1123–1126

Chen SM, Chen CD (2011) Handling forecasting problems based on high-order fuzzy logical relationships. Expert Syst Appl 38(4):3857–3864

## 

|  |  |  |
| --- | --- | --- |
| Name | ID | PART |
| MD. Sabbir Hasan | 17-33918-1 | Abstract & Conclusions |
| Mehedi Hasan Joy | 18-38413-2 | Introduction |
| MD. Samiul Hauque Chowdhury | 18-36072-1 | Literature Review |
| Shihab Hossain | 18-38131-2 | Results |
| Fahad, Abdullah Al | 18-37904-2 | Discussion |