I.J.Modern Education and Computer Science, **2013, 11, 11-20**
Published Online November 2013 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2013.11.02

Evaluation of Ensemble Classifiers for Handwriting Recognition

M.Govindarajan
Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar – 608002
Tamil Nadu, India.
Email: govind_aucse@yahoo.com

Abstract — One of the major developments in machine learning in the past decade is the ensemble method, which finds highly accurate classifier by combining many moderately accurate component classifiers. In this research work, new ensemble classification methods are proposed for homogeneous ensemble classifiers using bagging and heterogeneous ensemble classifiers using arcing classifier and their performances are analyzed in terms of accuracy. A Classifier ensemble is designed using Radial Basis Function (RBF) and Support Vector Machine (SVM) as base classifiers. The feasibility and the benefits of the proposed approaches are demonstrated by the means of real and benchmark data sets of recognizing totally unconstrained handwritten numerals. The main originality of the proposed approach is based on three main parts: preprocessing phase, classification phase and combining phase. A wide range of comparative experiments are conducted for real and benchmark data sets of recognizing totally unconstrained handwritten numerals. The accuracy of base classifiers is compared with homogeneous and heterogeneous models for data mining problem. The proposed ensemble methods provide significant improvement of accuracy compared to individual classifiers and also heterogeneous models exhibit better results than homogeneous models for real and benchmark data sets of recognizing totally unconstrained handwritten numerals.

Index Terms — Data Mining, Ensemble, Handwriting Recognition, Radial Basis Function, Support Vector Machine, Accuracy.

I. INTRODUCTION

Optical Character Recognition (OCR) is a branch of pattern recognition, and also a branch of computer vision. OCR has been extensively researched for more than four decades. With the advent of digital computers, many researchers and engineers have been engaged in this interesting topic. It is not only a newly developing topic due to many potential applications, such as bank check processing, postal mail sorting, automatic reading of tax forms and various handwritten and printed materials, but it is also a benchmark for testing and verifying new pattern recognition theories and algorithms. In recent years, many new classifiers and feature extraction algorithms have been proposed and tested on various OCR databases and these techniques have been used in wide applications. Numerous scientific papers and inventions in OCR have been reported in the literature. It can be said that OCR is one of the most important and active research fields in pattern recognition. Today, OCR research is addressing a diversified number of sophisticated problems. Important research in OCR includes degraded (heavy noise) omni font text recognition, and analysis/recognition of complex documents (including texts, images, charts, tables and video documents). Handwritten numeral recognition, (as there are varieties of handwriting styles depending on an applicant's age, gender, education, ethnic background, etc., as well as the writer's mood while writing), is a relatively difficult research field in OCR.

In the area of character recognition, the concept of combining multiple classifiers is proposed as a new direction for the development of highly reliable character recognition systems (C.Y.Suen et al., 1990) and some preliminary results have indicated that the combination of several complementary classifiers will improve the performance of individual classifiers (C.Y.Suen et al., 1990 and T.K.Ho et al., 1990). The primary objective of this paper is ensemble of radial basis function and Support Vector Machine is superior to individual approach for recognizing totally unconstrained handwritten numerals in terms of classification accuracy.

Data mining methods may be distinguished by either supervised or unsupervised learning methods. One of the most active areas of research in supervised learning has been to study methods for constructing good ensembles of classifiers. It has been observed that when certain classifiers are ensembled, the performance of the individual classifiers.

Recently, advances in knowledge extraction techniques have made it possible to transform various kinds of raw data into high level knowledge. However, the classification results of these techniques are affected by the limitations associated with individual techniques. Hence, hybrid approach is widely recognized by the data mining research community.

Hybrid models have been suggested to overcome the defects of using a single supervised learning method, such as radial basis function and support vector machine techniques. Hybrid models combine different methods to improve classification accuracy. The term combined model is usually used to refer to a concept similar to a

hybrid model. Combined models apply the same algorithm repeatedly through partitioning and weighting of a training data set. Combined models also have been called Ensembles. Ensemble improves classification performance by the combined use of two effects: reduction of errors due to bias and variance (Haykin, 1999).

This paper proposes new ensemble classification methods to improve the classification accuracy. The main purpose of this paper is to apply homogeneous and heterogeneous ensemble classifiers for real and benchmark dataset of recognizing totally unconstrained handwritten numerals to improve classification accuracy. Organization of this paper is as follows. Section 2 describes the related work. Section 3 presents proposed methodology and Section 4 explains the performance evaluation measures. Section 5 focuses on the experimental results and discussion. Finally, results are summarized and concluded in section 6.

II. RELATED WORK

In the past several decades, a wide variety of approaches have been proposed to attempt to achieve the recognition system of handwritten numerals. These approaches generally fall into two categories: statistical method and syntactic method (C. Y. Suen, et al., 1992). First category includes techniques such as template matching, measurements of density of points, moments, characteristic loci, and mathematical transforms. In the second category, efforts are aimed at capturing the essential shape features of numerals, generally from their skeletons or contours. Such features include loops, endpoints, junctions, arcs, concavities and convexities, and strokes.

Suen et al.,(1992) proposed four experts for the recognition of handwritten digits. In expert one, the skeleton of a character pattern was decomposed into branches. The pattern was then classified according to the features extracted from these branches. In expert two, a fast algorithm based on decision trees was used to process the more easily recognizable samples, and a relaxation process was applied to those samples that could not be uniquely classified in the first phase. In expert three, statistical data on the frequency of occurrence of features during training were stored in a database. This database was used to deduce the identification of an unknown sample. In expert four, structural features were extracted from the contours of the digits. A tree classifier was used for classification. The resulting multiple-expert system proved that the consensus of these methods tended to compensate for individual weakness, while preserving individual strengths. The high recognition rates were reported and compared favorably with the best performance in the field.

The utilization of the Support Vector Machine (SVM) classifier has gained immense popularity in the past years (C. J. C. Burges., et al., 1997 and U. Krebel, 1999). SVM is a discriminative classifier based on Vapnik's structural risk minimization principle. It can be implemented on flexible decision boundaries in high dimensional feature spaces. Generally, SVM solves a binary (two-class) classification problem, and multi-class classification is accomplished by combining multiple binary SVMs. Good results on handwritten numeral recognition by using SVMs can be found in Dong, et al.'s paper.

Renata F. P. Neves et al (2011) have proposed SVM based offline handwritten digit recognition. Authors claim that SVM outperforms the Multilayer perceptron classifier. Experiment is carried out on NIST SD19 standard dataset.Advantage of MLP is that it is able to segment non-linearly separable classes. However, MLP can easily fall into a region of local minimum, where the training will stop assuming it has achieved an optimal point in the error surface. Another hindrance is defining the best network architecture to solve the problem, considering the number of layers and the number of perceptron in each hidden layer. Because of these disadvantages, a digit recognizer using the MLP structure may not produce the desired low error rate

Muhammad et al (2012) have discussed hybrid feature extraction in their work. SVM is used as a classifier. Authors have combined structural, statistical and correlation functions to derive hybrid features. In first step, elementary stroke location is identified with the help of chosen elementary shape. To make it more robust, certain structural / statistical features are added in it. The added structural / statistical features are based on projections, profiles, invariant moments, endpoints and junction points. This enhanced, powerful combination of features results in a 157-variable feature vector for each character. It includes 100 correlation features and 57 structural/statistical features. Correlation features are based on Pearson's correlation coefficient.

Shubhangi et al, (2009) have extract similar correlation function based features for Chinese hand-printed character recognition. Classification is done based on minimum distance decision rule. While proposed method perform final classification based on support vector machine (SVM).

Artificial Neural Networks (ANN), due to its useful properties such as: highly parallel mechanism, excellent fault tolerance, adaptation, and self-learning, have become increasingly developed and successfully used in character recognition (A. Amin, et al., 1996 and J. Cai, et al., 1995). The key power provided by such networks is that they admit fairly simple algorithms where the form of nonlinearity that can be learned from the training data. The models are thus extremely powerful, have nice theoretical properties, and apply well to a vast array of real-world applications.

Malayalam is a language spoken by millions of people in the state of Kerala and the union territories of Lakshadweep and Pondicherry in India. It is written mostly in clockwise direction and consists of loops and curves. Neural network based approach is discussed in (Amritha Sampath et al, 2012) for Malayalam language. In pre processing step, noise is removed by applying

threshold (number of pixels in rectangular bounding box).

Postal address recognition system for Arabic language is proposed by M.Charfi et al. (2012) Writing translates style of writing, Mood and personality of the writer, which makes it difficult to characterize. From scanned envelop, printed boarder and stamp logo are suppressed. Address is located and using histogram method, lines, words and characters are segmented. Temporal order of strokes can be helpful for robust recognition. In literature, way of temporal order reconstruction is proposed. End stroke point, Branching point and Crossing point are detected from city name. Elliptical model is applied on preprocessed digit or character and matching process is applied.

Xu et al. (1992) proposed four combining classifier approaches according to the levels of information available from the various classifiers. The experimental results showed that the performance of individual classifiers could be improved significantly. Huang and Suen (1993, 1995) proposed the Behavior-Knowledge Space method in order to combine multiple classifiers for providing abstract level information for the recognition of handwritten numerals. Lam and Suen (1995) studied the performance of combination methods that were variations of the majority vote. A Bayesian formulation and a weighted majority vote (with weights obtained through a genetic algorithm) were implemented, and the combined performances of seven classifiers on a large set of handwritten numerals were analyzed.

Freund and Schapire (1995,1996) proposed an algorithm the basis of which is to adaptively resample and combine (hence the acronym--arcing) so that the weights in the resampling are increased for those cases most often misclassified and the combining is done by weighted voting.

Previous work has demonstrated that arcing classifiers is very effective for RBF-SVM hybrid system. (M.Govindarajan et al., 2012). A hybrid model can improve the performance of basic classifier (Tsai 2009).

In this paper, a hybrid handwriting recognition system is proposed using radial basis function and support vector machine and the effectiveness of the proposed bagged RBF, bagged SVM and RBF-SVM hybrid system is evaluated by conducting several experiments on real and benchmark datasets of handwriting recognition. The performance of the proposed bagged RBF, bagged SVM and RBF-SVM hybrid classifiers are examined in comparison with standalone RBF and standalone SVM classifier and also heterogeneous models exhibits better results than homogeneous models for real and benchmark data sets of recognizing totally unconstrained handwritten numerals.

III. PROPOSED METHODOLOGY

A. Preprocessing of real and benchmark datasets

The real dataset consists of images selected from the first 1000 images in the MNIST dataset. Weka provides a supervised instance filter named Resample that you can use to extract sample subsets of the MNIST dataset (the filter is "supervised" because it looks at the class labels in order to ensure that the class distribution is approximately the same in the samples as in the original dataset). In the Preprocessing tab, select Filters, then Supervised, then Instance, and finally resample. You can specify what percentage of the dataset should be used in a given sample. Generate and save samples containing 100, 250, and 500 instances, together with the full dataset for this assignment, which contains 1000 instances.

The benchmark data is related with Zip codes dataset. Locating the zip code on the envelope and separating each digit from its neighbors, a very hard task in itself, was performed by postal Service contractors (wang and Srihari 1998). At this point, the size of a digit image varies but is typically around 40 by 60 pixels. A linear transformation is then applied to make the image fit in a 16 by 16 pixel image. This transformation preserves the aspect ratio of the character, and is performed after extraneous marks in the image have been removed. Because of the linear transformation, the resulting image is not binary but has multiple gray levels, since a variable number of pixels in the original image can fall into a given pixel in the target image. The gray levels of each image are scaled and translated to fall within the rang -1 to 1.

B. Existing Classification Methods

1) Radial Basis Function Neural Network

Radial basis function (RBF) networks (Oliver Buchtala et al, 2005) combine a number of different concepts from approximation theory, clustering, and neural network theory. A key advantage of RBF networks for practitioners is the clear and understandable interpretation of the functionality of basis functions. Also, fuzzy rules may be extracted from RBF networks for deployment in an expert system.

The RBF networks used here may be defined as follows.

1. RBF networks have three layers of nodes: input layer uI , hidden layer u_H and output layer u_0
2. Feed-forward connections exist between input and hidden layers, between input and output layers (shortcut connections), and between hidden and output layers. Additionally, there are connections between a bias node and each output node. A scalar weight wi,j is associated with the connection between nodes i and j.
3. The activation of each input node (fanout) $i \in u_I$ is equal to its external input

$$a_i(k) \overset{def}{=} x_i(k) \tag{3.1}$$

where $x_i(k)$ is the element of the external input vector

(pattern) $X(k)$ of the network ($k = 1,2,....$ denotes the number of the pattern).

4. Each hidden node (neuron) $j \in u_H$ determines the Euclidean distance between "its own" weight vector $\mathrm{Wj} \overset{\text{def}}{=} (\mathrm{w}_{(1,\mathrm{j})},....., \mathrm{w}_{(|\mathrm{u_I}|,\mathrm{j})})^{\mathrm{T}}$ and the activations of the input nodes, i.e., the external input vector

$$\mathrm{s_j(k)} \overset{\text{def}}{=} \|\mathrm{Wj} - \mathrm{X(k)}\| \tag{3.2}$$

The distance $\mathrm{s_j(k)}$ is used as an input of a radial basis function in order to determine the activation $\mathrm{a_j(k)}$ of node j. Here, Gaussian functions are employed

$$a_j(k) \overset{def}{=} e^{(-s_j(k)2\,/\,r_j{}^2)} \tag{3.4}$$

The parameter $\mathrm{r_j}$ of node j is the radius of the basis function; the vector Wj is its center.
Localized basis functions such as the *Gaussian* or the *inverse multiquadric* are usually preferred.

5. Each output node (neuron) $l \in u_0$ computes its activation as a weighted sum

$$a_l(k) \overset{def}{=} \sum_{j=1}^{|u_H|} w_{(j,l)}.a_j(k) + \sum_{i=1}^{|u_I|} w_{(i,l)}.a_i(k) + w_{(B,l)} \tag{3.5}$$

The external output vector of the network $y(k)$ consists of the activations of output nodes, i.e. $y_1(k) \overset{def}{=} a_1(k)$. The activation of a hidden node is high if the current input vector of the network is "similar" (depending on the value of the radius) to the center of its basis function. The center of a basis function can, therefore, be regarded as a prototype of a hyperspherical cluster in the input space of the network. The radius of the cluster is given by the value of the radius parameter. In the literature, some variants of this network structure can be found, some of which do not contain shortcut connections or bias neurons.

2) Support Vector Machine

Support vector machines (Cherkassky et al., 1998; Burges, 1998) are powerful tools for data classification. Classification is achieved by a linear or nonlinear separating surface in the input space of the dataset. The separating surface depends only on a subset of the original data. This subset of data, which is all that is needed to generate the separating surface, constitutes the set of support vectors. In this study, a method is given for selecting as small a set of support vectors as possible which completely determines a separating plane classifier. In nonlinear classification problems, SVM tries to place a linear boundary between two different classes and adjust it in such a way that the margin is maximized (Vanajakshi and Rilett, 2004). Moreover, in the case of linearly separable data, the method is to find the most suitable one among the hyperplanes that minimize the training error. After that, the boundary is adjusted such that the distance between the boundary and the nearest data points in each class is maximal.

In a binary classification problem, its data points are given as:

$$D=\{(x^1,y^1),....(x^l,y^l)\},..x\in\Re^n,y\in\{-1,1\}, \tag{3.6}$$

where
y = a binary value representing the two classes and,
x = the input vector.

As mentioned above, there are numbers of hyperplanes that can separate these two sets of data and the problem is to find the hyperplane with the largest margin. Suppose that all training data satisfy the following constraints:

$$w.x + b \geq +1 \text{ for } y_i = +1 \tag{3.7}$$

$$w.x + b \leq -1 \text{ for } y_i = -1 \tag{3.8}$$

where
w = the boundary
x = the input vector
b = the scalar threshold (bias).

Therefore, the decision function that can classify the data is:

$$f(y) = \mathrm{sgn}((w.x) + b) \tag{3.9}$$

Thus, the separating hyperplane must satisfy the following constraints:

$$y_i[(w.x_i) + b] \geq 1 \tag{3.10}$$

where l = the number of training sets

The optimal hyperplane is the unique one that not only separates the data without error but also maximizes the margin. It means that it should maximize the distance between closest vectors in both classes to the hyperplane. Therefore the hyperplane that optimally separate the data into two classes can be shown to be the one that minimize the functional:

$$\phi(w) = \frac{|w|^2}{2} \tag{3.11}$$

Therefore, the optimization problem can be formulated into an equivalent non-constraint optimization problem by introducing the Lagrange multipliers ($\alpha_I >= 0$) and a Lagrangian:

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{t=1..l} \alpha t(yt((w.xt)+b)-1) \quad (3.12)$$

The Lagrangian has to be minimized with respect to **w** and *b* by the given expressions:

$$w_0 = \sum y\alpha x \quad (3.13)$$

This expressions for w0 is then substitute into equation (3.12) which will result in dual form of the function which has to be maximized with respect to the constraints $\alpha_I > 0$.
Maximize

$$W(\alpha) = \sum \alpha_I - \frac{1}{2} \sum_{I,j=1..l} \alpha_i \alpha_j y_i y_j (x_i x_j) \quad (3.14)$$

Subject to $\alpha_I \geq 0, i = 1..l$ and $\sum \alpha_I y_i$

The hyperplane decision function can therefore be written as:

$$f(x) = sign(w_0 x + b_0) = sign(\sum y_i \alpha_i^0 (x_i . x) + b_0) \quad (3.15)$$

However, the equation (3.15) is meant for linearly separable data in SVM. In a non-linearly separable data, SVM is used to learn the decision functions by first mapping the data to some higher dimensional feature space and constructing a separating hyperplane in this space.

C. Homogeneous Ensemble Classifiers using Bagging

1) Proposed Bagged RBF and SVM Classifiers

Given a set D, of d tuples, bagging (Breiman, L. 1996a) works as follows. For iteration i (i =1, 2,…..k), a training set, D_i, of d tuples is sampled with replacement from the original set of tuples, D. The bootstrap sample, D_i, created by sampling D with replacement from the given training data set D repeatedly. Each example in the given training set D may appear repeatedly or not at all in any particular replicate training data set D_i. A classifier model, M_i, is learnt for each training set, D_i. To classify an unknown tuple, X, each classifier, M_i, returns its class prediction, which counts as one vote. The bagged RBF and SVM, M^*, counts the votes and assigns the class with the most votes to X.

Algorithm: RBF and SVM ensemble classifiers using bagging
Input:

- D, a set of d tuples.
- *k* = 1, the number of models in the ensemble.
- Base Classifiers (Radial Basis Function, Support Vector Machine)

Output: Bagged RBF and SVM, M^*
Method:

1. for *i* = 1 to *k* do // create k models
2. Create a bootstrap sample, Di, by sampling D with replacement, from the given training data set D repeatedly. Each example in the given training set D may appear repeated times or not at all in any particular replicate training data set Di
3. Use Di to derive a model, Mi;
4. Classify each example d in training data Di and initialized the weight, Wi for the model, Mi, based on the accuracies of percentage of correctly classified example in training data Di.
5. endfor

To use the bagged RBF and SVM models on a tuple, X:

1. if classification then
2. let each of the k models classify X and return the majority vote;
3. if prediction then
4. let each of the k models predict a value for X and return the average predicted value;

D. Heterogeneous Ensemble Classifiers using Arcing

1) Proposed RBF-SVM Hybrid System

Given a set D, of d tuples, arcing (Breiman. L, 1996) works as follows; For iteration i (i =1, 2,…..k), a training set, D_i, of d tuples is sampled with replacement from the original set of tuples, D. some of the examples from the dataset *D* will occur more than once in the training dataset D_i. The examples that did not make it into the training dataset end up forming the test dataset. Then a classifier model, M_i, is learned for each training examples *d* from training dataset D_i. A classifier model, M_i, is learned for each training set, D_i. To classify an unknown tuple, X, each classifier, M_i, returns its class prediction, which counts as one vote. The hybrid classifier (RBF-SVM), M^*, counts the votes and assigns the class with the most votes to X.

Algorithm: Hybrid RBF-SVM using Arcing Classifier
Input:

- D, a set of d tuples.
- *k* = 2, the number of models in the ensemble.
- Base Classifiers (Radial Basis Function, Support Vector Machine)

Output: Hybrid RBF-SVM model, M^*.
Procedure:

1. For i = 1 to k do // Create k models
2. Create a new training dataset, Di, by sampling D with replacement. Same

example from given dataset D may occur more than once in the training dataset Di.
3. Use Di to derive a model, Mi
4. Classify each example d in training data Di and initialized the weight, Wi for the model, Mi, based on the accuracies of percentage of correctly classified example in training data Di.
5. endfor

To use the hybrid model on a tuple, X:

1. if classification then
2. let each of the k models classify X and return the majority vote;
3. if prediction then
4. let each of the k models predict a value for X and return the average predicted value;

The basic idea in Arcing is like bagging, but some of the original tuples of D may not be included in Di, where as others may occur more than once.

IV. PERFORMANCE EVALUATION MEASURES

A. Cross Validation Technique

Cross-validation (Jiawei Han and Micheline Kamber**,** 2003) sometimes called rotation estimation, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. 10-fold cross validation is commonly used. In stratified K-fold cross-validation, the folds are selected so that the mean response value is approximately equal in all the folds.

B. Criteria for Evaluation

The primary metric for evaluating classifier performance is classification Accuracy: the percentage of test samples that the ability of a given classifier to correctly predict the label of new or previously unseen data (i.e. tuples without class label information). Similarly, the accuracy of a predictor refers to how well a given predictor can guess the value of the predicted attribute for new or previously unseen data.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Real dataset Description

The dataset used to train and test the systems described in this paper was constructed from NIST's Special Database 3 and Special Database 1 which contain binary images of handwritten digits. NIST originally designated SD-3 as their training set and SD-1 as their test set. However, SD-3 is much cleaner and easier to recognize than SD-1. The reason for this can be found on the fact that SD-3 was collected among Census Bureau employees, while SD-1 was collected among high-school students. Drawing sensible conclusions from learning experiments requires that the result be independent of the choice of training set and test among the complete set of samples. Therefore it was necessary to build a new database by mixing NIST's datasets.

B. Benchmark dataset Description

The data used in classification is 10 % U.S. Zip code, which consists of selected records of the complete U.S. Zip code database. The database used to train and test the hybrid system consists of 4253 segmented numerals digitized from handwritten zip codes that appeared on U.S. mail passing through the Buffalo, NY post office. The digits were written by many different people, using a great variety of sizes, writing styles, and instruments, with widely varying amounts of care.

C. Experiments and Analysis

In this section, new ensemble classification methods are proposed for homogeneous ensemble classifiers using bagging and heterogeneous ensemble classifiers using arcing classifier and their performances are analyzed in terms of accuracy.

1) Homogeneous Ensemble Classifiers using Bagging

The NIST and U.S. Zip code datasets are taken to evaluate the proposed Bagged RBF and bagged SVM classifiers.

a) Proposed Bagged RBF and Bagged SVM

TABLE I. THE PERFORMANCE OF BASE AND PROPOSED BAGGED CLASSIFIERS FOR REAL DATASET

Real Dataset	Classifiers	Classification Accuracy
NIST dataset	RBF	76.5 %
	Proposed Bagged RBF	91.8 %
	SVM	89.2 %
	Proposed Bagged SVM	98.0 %

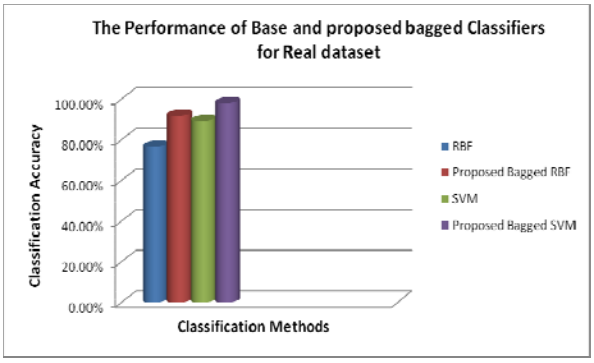

Figure 1. Classification Accuracy of Base and Proposed Bagged Classifiers Using Real dataset

TABLE II. THE PERFORMANCE OF BASE AND PROPOSED BAGGED CLASSIFIERS FOR BENCHMARK DATASET

Benchmark Dataset	Classifiers	Classification Accuracy
U.S. Zip code dataset	RBF	86.46 %
	Proposed Bagged RBF	97.74 %
	SVM	93.98 %
	Proposed Bagged SVM	95.45 %

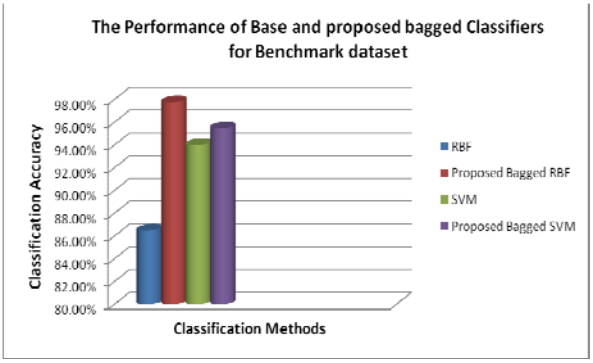

Figure 2. Classification Accuracy of Base and Proposed Bagged Classifiers Using Benchmark Dataset

In this research work, new ensemble classification methods are proposed for homogeneous ensemble classifiers using bagging and their performances are analyzed in terms of accuracy. Here, the base classifiers are constructed using radial basis function and Support Vector Machine. 10-fold cross validation (Kohavi, R, 1995) technique is applied to the base classifiers and evaluated Classification accuracy. Bagging is performed with radial basis function classifier and support vector machine to obtain a very good classification performance. Table 1 and 2 show classification performance for real and benchmark datasets of recognizing totally unconstrained handwritten numerals using existing and proposed bagged radial basis function neural network and support vector machine. The analysis of results shows that the proposed bagged radial basis function and bagged support vector machine classifies are shown to be superior to individual approaches for real and benchmark datasets of handwriting recognition problem in terms of classification accuracy. According to figure 1 and 2 proposed combined models show significantly larger improvement of Classification accuracy than the base classifiers. This means that the combined methods are more accurate than the individual methods in the field of handwriting recognition.

2) Heterogeneous Ensemble Classifiers using Arcing

The NIST and U.S. Zip code datasets are taken to evaluate the proposed hybrid RBF-SVM classifiers.

a) Proposed Hybrid RBF-SVM System

TABLE III. THE PERFORMANCE OF BASE AND PROPOSED HYBRID RBF-SVM CLASSIFIERS FOR REAL DATASET

Real Dataset	Classifiers	Classification Accuracy
NIST dataset	RBF	76.5 %
	SVM	89.2 %
	Proposed Hybrid RBF-SVM	99.3 %

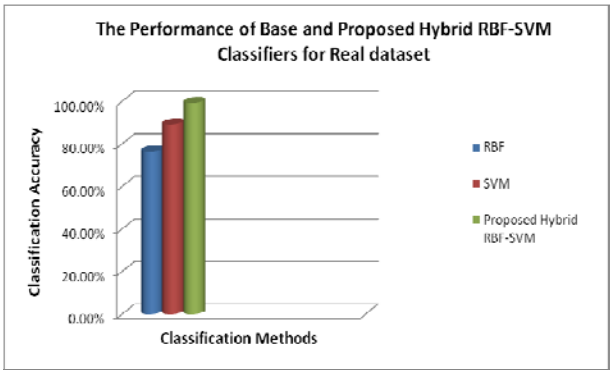

Figure 3. Classification Accuracy of Base and Proposed Hybrid RBF-SVM Classifiers Using Real Dataset

TABLE IV. THE PERFORMANCE OF BASE AND PROPOSED HYBRID RBF-SVM CLASSIFIER FOR BENCHMARK DATASET

Benchmark Dataset	Classifiers	Classification Accuracy
U.S. Zip code dataset	RBF	86.46 %
	SVM	93.98 %
	Proposed Hybrid RBF-SVM	99.13 %

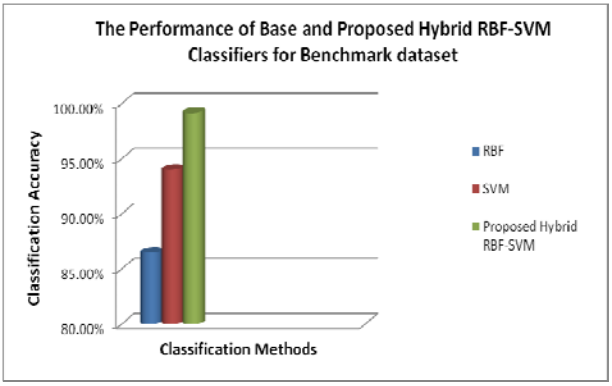

Figure 4. Classification Accuracy of Base and Proposed Hybrid RBF-SVM Classifiers Using Benchmark Dataset

In this research work, new hybrid classification methods are proposed for heterogeneous ensemble classifiers using arcing classifier and their performances are analyzed in terms of accuracy. The data set described in section 5 is being used to test the performance of base classifiers and hybrid classifier. Classification accuracy was evaluated using 10-fold cross validation. In the proposed approach, first the base classifiers RBF and SVM are constructed individually to obtain a very good generalization performance. Secondly, the ensemble of RBF and SVM is designed. In the ensemble approach, the final output is decided as follows: base classifier's output is given a weight (0–1 scale) depending on the generalization performance as given in Table 3 and 4. According to figure 3 and 4, the proposed hybrid models show significantly larger improvement of classification accuracy than the base classifiers and the results are found to be statistically significant.

The experimental results show that proposed hybrid RBF-SVM is superior to individual approaches for handwriting recognition problem in terms of classification accuracy.

VI. CONCLUTIONS

In this research work, new combined classification methods are proposed for in homogeneous ensemble classifiers using bagging and the performance comparisons have been demonstrated using real and benchmark dataset of handwriting recognition in terms of accuracy. Here, the proposed bagged radial basis function and bagged support vector machine combines the complementary features of the base classifiers. Similarly, new hybrid RBF-SVM models are designed in heterogeneous ensemble classifiers involving RBF and SVM models as base classifiers and their performances are analyzed in terms of accuracy

The experiment results lead to the following observations.

- SVM exhibits better performance than RBF in the important respects of accuracy.
- The proposed bagged methods are shown to be significantly higher improvement of classification accuracy than the base classifiers.
- The hybrid RBF-SVM shows higher percentage of classification accuracy than the base classifiers.
- The χ^2 statistic is determined for all the above approaches and their critical value is found to be less than 0.455. Hence corresponding probability is $p < 0.5$. This is smaller than the conventionally accepted significance level of 0.05 or 5%. Thus examining a χ^2 significance table, it is found that this value is significant with a degree of freedom of 1. In general, the result of χ^2 statistic analysis shows that the proposed classifiers are significant at $p < 0.05$ than the existing classifiers.
- The accuracy of base classifiers is compared with homogeneous and heterogeneous models for data mining problems and heterogeneous models exhibit better results than homogeneous models for real and benchmark data sets of handwriting recognition.
- The handwriting recognition dataset could be detected with high accuracy for homogeneous and heterogeneous models.

The future research will be directed towards developing more accurate base classifiers particularly for the handwriting recognition problem.

ACKNOWLEDGMENT

Author gratefully acknowledges the authorities of Annamalai University for the facilities offered and encouragement to carry out this work.

REFERENCES

[1] A. Amin, H. B. Al-Sadoun, and S. Fischer, "Hand-printed Arabic Character Recognition System Using An Artificial Network", Pattern Recognition Vol. 29, No. 4, 1996: 663-675.

[2] Amritha Sampath, Tripti C, Govindaru V, "Freeman code based online handwritten character recognition for Malayalam using backpropagation neural networks", International journal on Advanced computing, Vol. 3, No. 4, 2012: 51 – 58.

[3] Breiman. L, Bias, Variance, and Arcing Classifiers", Technical Report 460, Department of Statistics, University of California, Berkeley, CA, 1996.

[4] Breiman, L. Bagging predictors. Machine Learning, 24(2):1996a:123–140.

[5] C. J. C. Burges and B. Scholkopf, "Improving the Accuracy and Speed of Support vector Learning Machine", Advanced in Neural Information Processing Systems 9, MIT Press, Cambridge, MA, 1997: 375-381.

[6] Burges, C. J. C. "A tutorial on support vector machines for pattern recognition", Data Mining and Knowledge Discovery, 2(2):1998:121-167.

[7] J. Cai, M. Ahmadi, and M. Shridhar, "Recognition of Handwritten Numerals with Multiple Feature and Multi-stage Classifier", Pattern Recognition, VOL. 28, No. 2, 1995:153-160.

[8] Cherkassky, V. and Mulier, F. "Learning from Data - Concepts, Theory and Methods", John Wiley & Sons, New York, 1998.

[9] J. X. Dong, A. Krzyzak, and C.Y. Suen, "Fast SVM Training Algorithm with Decomposition on Very Large Datasets", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, No. 4, 2005: 603-618.

[10] Freund, Y. and Schapire, R. "A decision-theoretic generalization of on-line learning and an application to boosting", In proceedings of the Second European Conference on Computational Learning Theory, 1995: 23-37.

[11] Freund, Y. and Schapire, R. "Experiments with a new boosting algorithm", In Proceedings of the

Thirteenth International Conference on Machine Learning, 1996:148-156 Bari, Italy.
[12] M.Govindarajan, RM.Chandrasekaran, "Intrusion Detection using an Ensemble of Classification Methods", In Proceedings of International Conference on Machine Learning and Data Analysis, San Francisco, U.S.A, 24-26 October, 2012, pages 459-464.
[13] Haykin, S. "Neural networks: a comprehensive foundation" (second ed.), New Jersey: Prentice Hall, 1999.
[14] T.K.Ho, J.J.Hull, and S.N.Srihari, "Combination of Structural Classifiers", in Proc. IAPR Workshop Syntatic and Structural Pattern Recog., 1990: 123-137.
[15] Y. S. Huang and C. Y. Suen, "An Optimal Method of Combining Multiple Classifiers for Unconstrained Handwritten Numeral Recognition", Proceedings of 3rd International Workshop on Frontiers in Handwriting Recognition, 1993.
[16] Y. S. Huang and C. Y. Suen, "A Method of Combining Experts for the Recognition of Unconstrained Handwritten Numerals", IEEE Transactions on PAMI, Vol. 17, No. 1, 1995: 90-94.
[17] Jiawei Han, Micheline Kamber**,** "Data Mining – Concepts and Techniques", Elsevier Publications, 2003.
[18] Kohavi, R. "A study of cross-validation and bootstrap for accuracy estimation and model selection", Proceedings of International Joint Conference on Artificial Intelligence, 1995:1137–1143.
[19] U. Krebel, "Pairwise Classification and Support Vector Machines, Advances in Kernel Methods: Support Vector Learning", MIT Press, Cambridge, MA, 1999: 255-268.
[20] L. Lam and C. Y. Suen, "Optimal Combinations of Pattern Classifiers", Pattern Recognition Letters, Vol. 16, No. 9, 1995: 945-954.
[21] Moncef Charfi, Monji Kherallah, Abdelkarim El Baati, Adel M. Alimi, "A New Approach for Arabic Handwritten Postal Addresses Recognition", International Journal of Advanced Computer Science and Applications, Vol. 3, No. 3, 2012: 1-7.
[22] Muhammad Naeem Ayyaz, Imran Javed, Waqar Mahmood, "Handwritten Character Recognition Using Multiclass SVM Classification with Hybrid Feature Extraction", Pakistan journal of Engineering and Application Science, Vol. 10, 2012: 57-67.
[23] Oliver Buchtala, Manuel Klimek, and Bernhard Sick, Member, IEEE, "Evolutionary Optimization of Radial Basis Function Classifiers for Data Mining Applications", IEEE Transactions on systems, man, and cybernetics—part b: cybernetics, vol. 35, no. 5, 2005.
[24] Renata F. P. Neves, Alberto N. G. Lopes Filho, Carlos A.B.Mello, CleberZanchettin, "A SVM Based Off-Line Handwritten Digit Recognizer", International conference on Systems, Man and Cybernetics, IEEE Xplore, 2011: 510-515, Brazil.
[25] D. C. Shubhangi and P. S. Hiremath, "Handwritten English character and digit recognition using multiclass SVM classifier and using structural micro features," International Journal of Recent Trends in Engineering, vol. 2, no. 2 2009.
[26] C.Y.Suen, C.Nadal, T.A.Mai, R.Legault, and L.Lam, "Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts," Frontiers in Handwriting Recognition , C.Y.Suen, Ed., IN Proc.Int.Workshop on Frontiers in Handwriting Recognition, Montreal, Canada, Apr. 2-3, 1990" 131-143.
[27] C. Y. Suen, C. Nadal, R. Legault, T. A. Mai, and L. Lam, (1992), "Computer recognition of unconstrained handwritten numerals," *Proc. IEEE*, vol. 80, 1992: 1162–1180.
[28] Tsai, C. F., Lu, Y.F. "Customer Churn Prediction by Hybrid Neural Network", Expert Systems with Application (39): 2009: 12547-12553.
[29] Vanajakshi, L. and Rilett, L.R. "A Comparison of the Performance of Artificial Neural Network and Support Vector Machines for the Prediction of Traffic Speed", IEEE Intelligent Vehicles Symposium, University of Parma, Parma, Italy: IEEE: 2004:194-199.
[30] Vapnik, V. Statistical learning theory, New York, John Wiley & Sons, 1998.
[31] Wang, C.H, and Srihari, S.N. "A framework for object recognition in a visually complex environment and its applications to locating address blocks on mail pieces", Int J Computer Vision 2, 125, 1998.
[32] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of Combining Multiple Classifiers and Their Applications to Handwritten Recognition", IEEE Transactions on Systems, Man, Cybernetics, Vol. 22, No. 3, 1992: 418-435.

M.Govindarajan received the B.E and M.E and Ph.D Degree in Computer Science and Engineering from Annamalai University, Tamil Nadu, India in 2001 and 2005 and 2010 respectively. He did his post-doctoral research in the Department of Computing, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey, United Kingdom in 2011 and pursuing Doctor of Science at Utkal University, orissa, India. He is currently an Assistant Professor at the Department of Computer Science and Engineering, Annamalai University, Tamil Nadu, India. He has presented and published more than 75 papers at Conferences and Journals and also received best paper awards. He has delivered invited talks at various national and international conferences. His current Research Interests include Data Mining and its applications, Web Mining, Text Mining, and Sentiment Mining. He was the recipient of the Achievement Award for the field and to

the Conference Bio-Engineering, Computer science, Knowledge Mining (2006), Prague, Czech Republic, Career Award for Young Teachers (2006), All India Council for Technical Education, New Delhi, India and Young Scientist International Travel Award (2012), Department of Science and Technology, Government of India New Delhi. He is Young Scientists awardee under Fast Track Scheme (2013), Department of Science and Technology, Government of India, New Delhi and also granted Young Scientist Fellowship (2013)**,** Tamil Nadu State Council for Science and Technology, Government of Tamil Nadu, Chennai. He has visited countries like Czech Republic, Austria, Thailand, United Kingdom, Malaysia, U.S.A, and Singapore. He is an active Member of various professional bodies and Editorial Board Member of various conferences and journals.

