Character Recognition using Ensemble classifier

Sadhana Tiwari
Information Technology Department
Galgotia College of Engineering and Technolgoy
Greater Noida, India
Sadhanatiwari.22@gmail.com

Abstract—To improve the accuracy of data classification systems, several techniques using classifier fusion have been suggested. This paper proposed a model of classifier fusion for character recognition problem. The work presented here aims to tackle the disadvantages and benefit of different classifiers with varying feature sets. In particular, this approach proposes the use of statistical procedures for the selection of the best subgroup among different classification algorithms and the subsequent fusion of the decision of the models in this subgroup with methods like voting, weighted voting. Ensemble classifier is constructed by using Support Vector Machine and K-Nearest Neighbor. Experimental results show that the performance of proposed ensemble classifier is better as compared to other classifiers in character recognition.

Keywords-Character Recognition; Ensemble classifier; Support Vector Machine; K-Nearest Neighbor.

I. INTRODUCTION

Data Mining plays an important role for finding useful information from given data. It is one of the widely accepted approach suitable for every field. The application area of data mining is very wide and includes several application such as medical diagnosis, fraud detection, image classification, emotion detection, rain forecasting, education, target detection, profitability prediction etc. [1-20]. Character recognition is also one of the vital application of data mining [21-23]. But the main issue of consideration is the selection of suitable data mining technique for character recognition. Classification, association, clustering, regression, outlier detection are some data mining techniques. For character recognition, we choose classification technique which includes several approaches such as Support Vector Machine, Neural Network, Decision tree, Naive Bayes, K-Nearest Neighbor etc. As we know, all classification approaches have their advantages and disadvantages. No classification technique is perfect in every situation. Classifier ensemble is one of the emerging and useful approach given by various researchers which combines the advantages of several classification techniques [24-31].

So, the main purpose of this research work is to improve the accuracy of data classification by fusion of classifiers with varying feature sets. The data classification problem used for performing the experiment is of Character Recognition. The dataset has been divided into two sets: the first set of data is used for training purpose and the second dataset is used for the testing of the model. In the first phase, pre-processing on image is done and after this features are extracted from the input image. The classifiers are then trained with the training data. We, right now, are considering 4 (four) classifiers (2 (two) distinct classifiers with 2(two) distinct feature sets). The performance of each classifier is evaluated by using confusion matrix in the second phase. In the third phase, a decision is inferred by combining the outputs of each classifier with considering confusion matrix which tells the confidence of each classifier in their results. The fusion approach used here uses the notion that different classifier when trained with the training data have different error rates in predicting output class. We, therefore, uses a 3 (three) step approach, instead of traditional 2 (two) step classification approach for data classification. There are generally two types of combination: classifier selection and classifier fusion. The presumption in classifier selection is that each classifier is an expert in some local area of the feature space. When a feature vector x is submitted for classification, the classifier responsible for the vicinity of x is given the highest credit when assigning the class label to x. We can nominate exactly one classifier to make the decision, or more than one local expert. Ensemble of classifiers assumes that all classifiers are trained over the whole feature space, and are thereby considered as competitive rather than complementary. Let $D = \{D_1, \ldots, D_L\}$ be a set (pool/committee/ensemble/team) of classifiers. By combining the individual outputs, we aim at a higher accuracy than that of the best classifier. There is a consensus among the researchers in classifier combination that the major factor for a better accuracy is the diversity in the classifier team and, so, the ensemble method is of a secondary importance. However, a choice of an appropriate fusion method can improve further on the performance of the ensemble. Some fusion methods are the most obvious choice when constructing a multiple classifier system i.e, the support for class w_i, $d_i(x)$, yielded by the team is, $d_i(x)=F(d_{1,i(x)},d_{2,i(x)},\ldots d_{l,i(x)})$ where i=1, 2, 3… and F= is the chosen Fusion method. Some choices for F are: Maximum, Minimum, Average, Median , Majority Vote etc.

This paper is divided into 4 sections. Ensemble classifier is discussed in section 2. Section 3 and 4 include proposed methodology and result of experiment respectively. Section 5 contains concluding remarks.

II. ENSEMBLE CLASSIFIER

Ensemble classifier is constructed by using Support Vector Machine and K-nearest neighbor which are discussed below:

A. Support Vector Machine

Generally, a SVM is a maximum margin classifier. The objective of the SVM optimization problem is to obtain certain parameters in order to define a separating hyper plane that has an optimal class separability (optimal in terms of maximum margin that is defined by the support vectors) [32-33].If the data is not linearly separable, the so-called Kernel-trick comes into play. There are different kinds of kernels such as Polynomial, Gaussian and Sigmoid kernel functions. In this research work, we have used linear kernel function. Kernels transform the original data into a higher dimensional feature space. Even if the original data are nonlinear, the transformed data is separable by a hyper-plane in feature space. Originally, Support Vector Machine is a binary classifier. So, for the application of character recognition, we use "One-versus-all" SVM with "*winner takes all"* strategy for multiclass classification [30]. For M-class classification problem, "One-versus-All" SVM classifier constructs *M* binary SVM classifiers. The output function p_i *of* i^{th} classifier consider the examples of i^{th} class as positive and the examples from all other classes as negative. For a new example **x**, "One-versus-All" SVM classifier assigns it to the class with the largest value of p_i.

B. K-Nearest Neighbor

The K-Nearest Neighbor's algorithm is a method for classifying objects based on the closest training examples in the feature space. KNN is a lazy learning algorithm. Even it is a very simple algorithm for it gives good results some of the time. KNN classification divides data into a training dataset and test dataset. For each row of the test set, the K nearest (in Euclidean distance) training set objects are found, and classification is determined by majority vote with ties broken at random.

Algorithm:

- Using the above features assign the training to data to its corresponding classes.
- For an input test image calculate its Euclidean distance with all the training data examples
- Now sort all the distances and take the k training examples with smallest distance and now we will decide the class of this test input according to the majority vote and if there is tie then we can take any value.

For deciding the value of k, we checked the algorithm for different values of k from 1 to 20 and then taken the most suitable value. K=4 is taken as the final value for KNN algorithm.

Classifiers Used:

Classifier 1: SVM with Feature set 1.

Classifier 2: KNN with Feature set 1.

Classifier 3: KNN with Feature set 2.

III. PROPOSED APPROACH

Figure 1 shows the proposed methodology for character recognition. The proposed model mainly includes three phases.

A. Pre-processing in an Image

Pre-processing an image is the first phase of the proposed approach which is done by the following ways:

- Convert an image into the gray scale (if the input image is colored).
- Convert this image to the binary image.
- Apply median filter on the image for removing the noise and then take the complement of this image.
- Do the dilation of the image. Dilation adds pixels to the boundaries of the objects in an image. The number of pixels added depends upon the size and shape of the structuring element used to process the image.

B. Feature Extraction

Feature extraction is the second phase of proposed methodology. We are considering two types of feature sets.

Set-1

In the structural features we are making a matrix of 2*10 dimensions means total of 20 features for each image, now how we are calculating this:

1.) First of all we scale our input image to 32*32(we are resizing every image to 32*32) and binarizing that image.

2.) Now we find the connected components of the image and assigning a value greater than equal to zero for each connected components

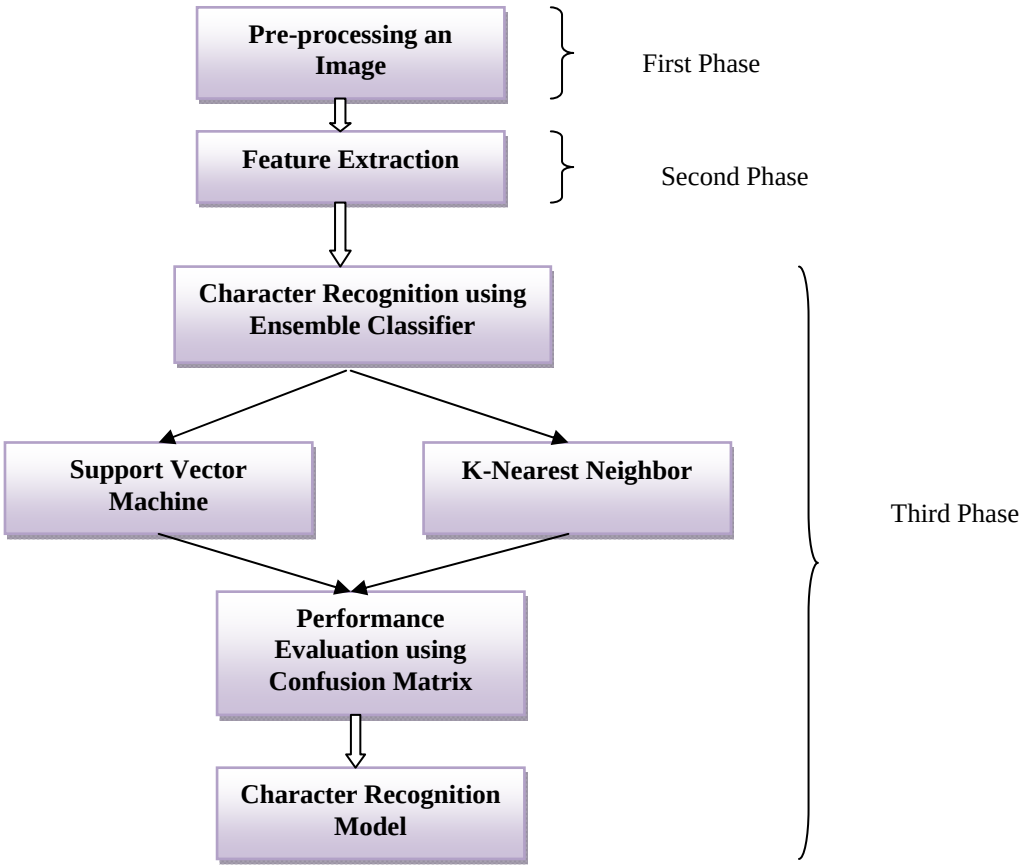

Figure 1. Proposed Model

0 value for the background pixels
1 for the first connected component
2 for the second connected component and so on. For example, input matrix is shown in figure 2 (a) and output label matrix is shown in figure 2(b):

1	0	1	1
1	0	1	1
1	0	1	1
1	0	1	1

Figure 2(a) Input Matrix

1	0	2	2
1	0	2	2
1	0	2	2
1	0	2	2

Figure 2(b) Output Label Matrix

3.) Now we are finding a sub-image that acts as a descriptor for the whole image example: Bounding box. We are finding the bounding box from the label matrix, bounding box gives us four values Starting pixel coordinates(x, y) and length, breadth. It is calculated as shown in figure 3:

1	0	0	1
1	1	1	1
0	0	1	1

Figure 3. Labeled Matrix

We need 4 points, starting position(x, y), length and breadth.

- Minimum value of row and column minus 0.5 gives starting position(x,y) respectively.
- Minimum value of row=1-0.5=0.5.
- Minimum value of column=1-0.5=0.5.

- Maximum value of column – minimum value of column+1 gives breadth of the box.
- Maximum value of column=4.
- Max value-min value of column=3+1.
- Maximum value of row- minimum value of row +1 gives length of the box.
- Maximum value of row=3.
- Max value – Min value=2+1.

4.) Bounding Box value for the given example: [0.5000 0.5000 4.0000 3.0000].
5.) Now we crop the image using these bounding box values.
6.) Now we are resizing the pic to 100*100 and now dividing it to two parts . First with row numbers 1 to 50 and then 51 to 100.

And taking 10 columns and then calculating the number of 'on' pixels for each subpart and taking this as one feature similarly calculating this for all 20 parts so a feature vector of 2*10 is formed.

Set-2 Fringe Map

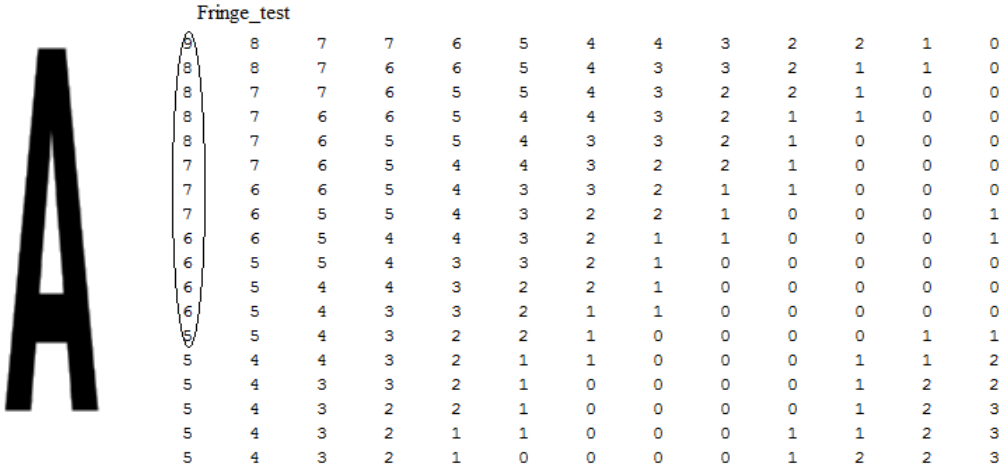

Figure 4. (a) English Character (b) Fringe map of image (a) and Peak Fringe Number in circle

The concept of fringe maps is related to distance transform for binary images. In a fringe map each pixel is represented with a fringe number. Every print (black) pixel has a fringe number of zero. Background (white) pixels have a fringe number which is a positive integer that is the distance from the nearest black pixel. In other words, a white pixel with fringe number *x* states that:

1) It is *x* pixels away from its nearest black pixel.

2) It is surrounded by at least (*x-1*) white pixels in all directions.

For us the second point is very useful. It helps us to quantify the white spacing between the components in the image. Assuming input to be normal binary images where the printing or writing is dark and background is light, so black pixels or the writing is represented as '0' and white pixels are set to a value '-1'. To generate a fringe map for the input binary image we start by examining each neighbor of each black pixel, and write a value of '1' into each neighbor which is a white pixel (-1). Horizontal, vertical, and diagonal neighbors are examined. Having done this, we examine each neighbor of each pixel with a '1' in it, and write '2' in the neighbors that are '-1'. We continue growing fringes and incrementing fringe counts until there are no more '-1' value pixels. For example we take the Tibetans character Figure 4(a) and its fringe map is shown in Figure 4 (b).

C. Performance Evaluation Parameter

Performance of ensemble classifier is evaluated using accuracy and confusion matrix. Let x_1, x_2,…,x_n represents training dataset. Now we define *Confusion Matrix (CM)* as a 2-D matrix of n*n (n is the total number of classes). Let C be a Confusion Matrix as shown in table 1. where C (i, j): When Class j is tested, it gives how many times the classifier called it to class i. So if we look at the Confusion Matrix Column, It states that that the classifier has detected the class correctly is given by C (i, i) i.e. the diagonal elements. Now we define,

TABLE I. CONFUSION MATRIX

Predicted Class ↓	Actual Class →			
	C-1	**C-2**	**...**	**C-n**
C-1	C(1,1)	C(1,2)		C(1,n)
C-2	C(2,1)	C(2,2)		C(2,n)
...				
C-n	C(n,1)	C(n,2)	...	C(n,n)

Where C(i,i) indicates total number of samples recognized correctly or total number of correct prediction made by the classifier for class i. Confidence P_i of a classifier for class 'i' is obtained as:

$$P_i = C(i, i)/\sum_{i=1 \text{ to } n} C(j,i) \quad (1)$$

For example, for class 1, the confidence of the classifier is defined as:

$$p_1(k) = \frac{C(1,1)}{C(1,1) + C(1,2) + C(1,3) + C(1,4) + C(1,5)} \quad (2)$$

Now if we have N classifiers, Let $Cl_1, Cl_2, \ldots Cl_N$ denotes N classifiers. The output of each classifier as o_1, $o_2 \ldots o_m$. Then we compute the total confidence for each class as given by, For Cl_k $Confidence_k = \sum_{j=1 \text{ to } m} Po_k(Cl_j)$ when classifier Cl_k has given the output as class k. Now we finally output the class with maximum Confidence Value. In this paper, we combine Support Vector Machine and K-nearest Neighbor due to their effective performance. Accuracy of a classifier is obtained as:

$$A = \frac{\sum_{i=1}^{n} C(i,i)}{\sum_{i=1}^{n} \sum_{j=1}^{n} C(i,j)} \quad (3)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This research work have used Tibetan Character set for the application of character recognition. There are 54 classes and the dataset has been divided into three sets. The first set (containing 50 images for each class) is used for the training purpose, second data set(containing 15 images for each class) is used for building up the confusion matrix for each classifier and the third data(containing 10 images for each class) set will be used for the purpose of testing. The images in the dataset are in PNG (Portable Network Graphics Format). We have performed experiment using Matlab. Table 2 and figure 5 shows the performance comparison of different classifiers for character recognition:

TABLE II. PERFORMANCE COMPARISON

Classifiers	**Accuracy**
KNN with Fringe Map	93.33%
KNN with structural feature set	95.69%
SVM with structural feature set	93.33%
Ensemble Classifier	**98.84%**

As shown by the results that the accuracy has increased in the approach suggested so the objective has achieved.

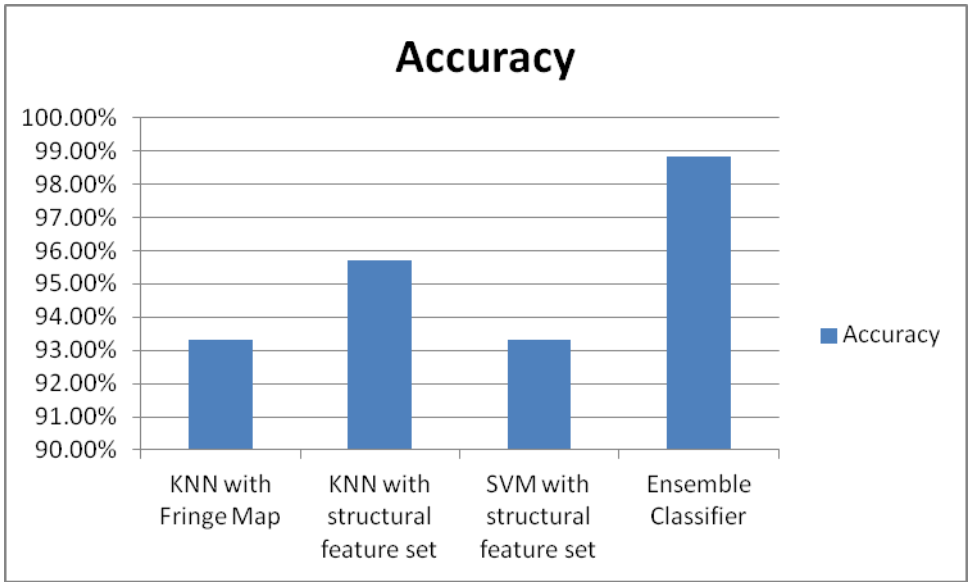

Figure 5 Accuracy Comparison of different Classifiers

V. CONCLUSION

This paper proposed an ensemble classifier for the character recognition. Fusion of Support Vector Machine and K-NN are used for the construction of ensemble classifier. The main purpose of this research work is to handle the disadvantages of classifiers and utilizes their strength. The proposed classifier uses statistical procedures for the selection of the best subgroup among different classification algorithms and the subsequent fusion of the decision of the models in this subgroup is performed by using weighted voting approach. It is analyzed form the experimental work that the performance of proposed ensemble classifier for character recognition dataset is better as compared to the SVM and KNN.

REFERENCES

[1] D. Tomar and S.. "A survey on Data Mining approaches for Healthcare." International Journal of Bio-Science & Bio-Technology 5, no. 5 (2013).
[2] D.W. Abbott, "Comparison of data analysis and classification algorithms for automatic target recognition", in IEEE International Conference on Systems, Man and Cybernetics, San Antonio, 1994.
[3] D. Tomar, D.Ojha, and S. Agarwal. "An Emotion Detection System Based on Multi Least Squares Twin Support Vector Machine." Advances in Artificial Intelligence 2014 (2014).
[4] S. Nath, S. and Q. A. Kazmi. "Image histogram segmentation by multi-level thresholding using Hill climbing algorithm." International Journal of Computer Applications 35, no. 1 (2011).
[5] D. Tomar and S. Agarwal. "Hybrid Feature Selection based Weighted Least Squares Twin Support Vector Machine approach for diagnosing Breast Cancer, Hepatitis and Diabetes." , Advances in Artificial Neural Systems, (2014).
[6] Khanna, Subham, and Sonali Agarwal. "An Integrated Approach towards the prediction of Likelihood of Diabetes." International Conference on Machine Intelligence Research and Advancement, 2013.
[7] D. Tomar, R. Arya, and S. Agarwal. "Prediction of profitability of industries using weighted SVR." International Journal on Computer Science and Engineering 3, no. 5 (2011): 1938-1945.
[8] M. Shukla and S. Agarwal. "Hybrid approach for tuberculosis data classification using optimal centroid selection based clustering." In Engineering and Systems (SCES), 2014 Students Conference on, pp. 1-5. IEEE, 2014.
[9] S. Agarwal and D. Tomar. "A Feature Selection Based Model for Software Defect Prediction." International Journal of Advanced Science and Technology, Vol. 65, pp.39-58, (2014).
[10] Y. Zhen, X. Chen, and P. Guo,Software Defect Prediction Using Fuzzy Support Vector Regression", Springer-Verlag Berlin Heidelberg, 2010.
[11] S. Agarwal and D.Tomar. "Prediction of software defects using Twin Support Vector Machine." In Information Systems and Computer Networks (ISCON), 2014 International Conference on, pp. 128-132. IEEE, 2014.
[12] S. Agarwal, "Weighted support vector regression approach for remote healthcare monitoring." In Recent Trends in Information Technology (ICRTIT), 2011 International Conference on, pp. 969-974. IEEE, 2011.
[13] Divya and S. Agarwal. "Classification of Countries based on Macro-Economic Variables using Fuzzy Support Vector Machine." International Journal of Computer Applications 27, no. 6 (2011).
[14] M. Kumar, A. Sharma, and S. Agarwal. "Clinical decision support system for diabetes disease diagnosis using optimized neural network." In Engineering and Systems (SCES), 2014 Students Conference on, pp. 1-6. IEEE, 2014.
[15] D. Tomar and S. Agarwal. "Predictive Model for diabetic patients using Hybrid Twin Support Vector Machine." 5th International Conferences on advances in communication Network and Computing (CNC-2014).
[16] A.K.Yadav, D. Tomar, and S. Agarwal. "Clustering of lung cancer data using Foggy K-means." In Recent Trends in Information Technology (ICRTIT), 2013 International Conference on, pp. 13-18. IEEE, 2013.
[17] D. Tomar and S. Agarwal. "Feature Selection based Least Square Twin Support Vector Machine for Diagnosis of Heart Disease." International Journal of Bio-Science & Bio-Technology 6, no. 2 (2014).
[18] S. Agarwal and S. K. Sinha. "Data mining based pervasive system design for Intensive Care Unit." In Computer Communication and Informatics (ICCCI), 2014 International Conference on, pp. 1-6. IEEE, 2014.

[19] D. Tomar, S. Singhal, and S. Agarwal. "Weighted Least Square Twin Support Vector Machine for Imbalanced Dataset." International Journal of Database Theory & Application 7, no. 2 (2014).
[20] S. Agarwal, G.N. Pandey, and M. Tiwari. "Data mining in education: data classification and decision tree approach." International Journal of e-Education, e-Business, e-Management and e-Learning, 2 (2) (2012): 140-144.
[21] V. Patil and S. Shimpi. "Handwritten English character recognition using neural network." Elixir Comput Sci Eng 41 (2011): 5587-5591.
[22] O. Matan, R. K. Kiang, C. E. Stenard, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, and Y. Le Cun. "Handwritten character recognition using neural network architectures." In Proceedings of the 4th United Postal Service Advanced Technology Conference, vol. 2, pp. 1003-1011. 1990.
[23] D.X. Ni. "Application of neural networks to character recognition." Proceedings of students/faculty research day, CSIS, Pace University, May 4th (2007).
[24] S. Nagi, D.K. Bhattacharyya, "Classification of microarray cancer data using ensemble approach" in Network Modeling Analysis in Health Informatics and Bioinformatics Volume 2, Issue 3, 2013, pp 159-173.
[25] N. Rathore and S. Agarwal. "Predicting the survivability of breast cancer patients using ensemble approach." In Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on, pp. 459-464. IEEE, 2014.
[26] T.G.Dietterich, "Ensemble methods in machine learning." In Multiple classifier systems, pp. 1-15. Springer Berlin Heidelberg, 2000.
[27] K.B. Duan and S. S. Keerthi,"Which is the Best Multiclass-SVM Method, An Empirical Study", Bioinformatics Research Centre,Nanyang Technological University, Nanyang Avenue, Singapore 639798.
[28] S. Agarwal and G. N. Pandey. "SVM based context awareness using body area sensor network for pervasive healthcare monitoring." In Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia, pp. 271-278. ACM, 2010.
[29] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman. "An ensemble classification-based approach applied to retinal blood vessel segmentation." Biomedical Engineering, IEEE Transactions on 59, no. 9 (2012): 2538-2548.
[30] V. Nikulin, J. McLachlan, and S. K. Ng. "Ensemble approach for the classification of imbalanced data." In AI 2009: Advances in Artificial Intelligence, pp. 291-300. Springer Berlin Heidelberg, 2009.
[31] L. I.Kuncheva,Member,IEEE,"A Theoretical Study on Six Classifier Fusion Strategies",IEEE transactions on Pattern Analysis And Machine Intelligence,vol. 24,No. 2,February 2002.
[32] C. Cortes , V. Vapnik , Support-vector network. Mach Learn 20:273–297, 1995.
[33] V. Vapnik, The nature of statistical learning, 2nd edn. Springer, New York, 1998.
[34] C.W. Hsu and C.J. Lin, A comparison of methods for multiclass support vector machines. Neural Networks, IEEE Transactions on, 13(2), 415-425, 2002.