A NOVEL FEATURE SET FOR RECOGNITION OF SIMILAR SHAPED HANDWRITTEN HINDI CHARACTERS USING MACHINE LEARNING

Sheetal Dabra[1], Sunil Agrawal[2] and Rama Krishna Challa[3]

[1]Department of Computer Science and Engineering, N.I.T.T.T.R, Chandigarh, India
er.sheetaldabra@gmail.com
[2]Department of Electronics and Communication, U.I.E.T, Punjab University, Chandigarh, India
s.agrawal@hotmail.com
[3]Department of Computer Science and Engineering,N.I.T.T.T.R,Chandigarh, India
rkc_97@yahoo.com

ABSTRACT

The growing need of handwritten Hindi character recognition in Indian offices such as passport, railway etc, has made it a vital area of research. Similar shaped characters are more prone to misclassification. In this paper four Machine Learning (ML) algorithms namely Bayesian Network, Radial Basis Function Network (RBFN), Multilayer Perceptron (MLP), and C4.5 Decision Tree are used for recognition of Similar Shaped Handwritten Hindi Characters (SSHHC) and their performance is compared. A novel feature set of 85 features is generated on the basis of character geometry. Due to the high dimensionality of feature vector, the classifiers can be computationally complex. So, its dimensionality is reduced to 11 and 4 using Correlation-Based (CFS) and Consistency-Based (CON) feature selection techniques respectively. Experimental results show that Bayesian Network is a better choice when used with CFS while C4.5 gives better performance with CON features.

KEYWORDS

Character Recognition, Feature Extraction, Machine Learning, Feature Selection

1. INTRODUCTION

In recent years, handwritten Hindi character recognition has grabbed a lot of attention as Hindi, being second official language in India, has wide applications in areas like passport, railway, postal address reading etc. Handwritten Character Recognition (HCR) consists of four main steps. First, the image samples of the handwritten characters are scanned using a scanner or digital camera. Pre-processing is done to improve the quality of the scanned images. Then features are extracted to form a feature vector which is input to the classifier for recognition.

Earlier, traditional classifiers such as Nearest Neighbor (NN), Hidden Markov Models (HMM) etc. were adopted for character recognition, however they exhibit certain limitations. Machine learning (ML) [1] algorithms provide a promising alternative in character recognition based on the feature set given to them. Each character image sample can be expressed in terms of some

D.C. Wyld, et al. (Eds): CCSEA 2011, CS & IT 02, pp. 25–35, 2011.
 DOI : 10.5121/csit.2011.1204

quantifiable attributes called features. A variety of features can be extracted such as primitives [2], profiles [3] etc. ML algorithm is then trained with this list of measured features, so that it maps these input features onto a class among certain predefined classes. Then the classifier can be used to determine the class of unknown samples used for testing.

Most of the misclassification by a classifier is due to similar shaped characters. Similarity in their structures leads to similar features, resulting in difficult for a machine to identify them correctly. Therefore, extensive research efforts are still needed to improve the recognition accuracy of similar shaped characters.

In this paper we evaluate and compare the performance of four ML algorithms namely Bayesian Network, Radial Basis Function Network (RBFN), Multilayer Perceptron (MLP), and C4.5Decision Tree for recognition of Similar Shaped Handwritten Hindi Characters (SSHHC). Feature set for the classifiers is generated on the basis of character geometry and 85 features are obtained. The size of this feature vector, being large in number, can lead to computational complexity of the classifiers. Therefore, we reduce these features using Correlation-Based (CFS) and Consistency-Based (CON) feature reduction techniques. Making one similar shaped handwritten Hindi character pair as the target pair for testing, the performance of the chosen ML algorithms is analyzed based on the following:

- Training the classifiers with a dataset consisting of few samples of only the target pair.
- Increasing the number of samples of the target pair in the training dataset.
- Further adding other similar shaped characters to this training dataset.
- Reducing the feature set with CFS and CON feature reduction algorithms.

The organization of the paper is as follows: Section 2 describes the related work. Section 3 illustrates the machine learning concepts with brief description of ML algorithms. Section 4 discusses experimental methodology with pre-processing steps, feature extraction technique and dataset used. Results and discussion is described in Section 5. Finally, conclusion and future work is presented in Section 6.

2. RELATED WORK

Character recognition task has been attempted through many different approaches like template matching, statistical techniques like NN, HMM, Quadratic Discriminant function (QDF) etc. Template matching works effectively for recognition of standard fonts, but gives poor performance with handwritten characters and when the size of dataset grows. It is not an effective technique if there is font discrepancy [4]. HMM models achieved great success in the field of speech recognition in past decades, however developing a 2-D HMM model for character recognition is found difficult and complex [5]. NN is found very computationally expensive in recognition purpose [6]. N. Araki et al. [7] applied Bayesian filters based on Bayes Theorm for handwritten character recognition. Later, discriminative classifiers such as Artificial Neural Network (ANN) and Support Vector Machine (SVM) grabbed a lot of attention. In [3] G. Vamvakas et al. compared the performance of three classifiers: Naive Bayes, K-NN and SVM and attained best performance with SVM. However SVM suffers from limitation of selection of kernel. ANNs can adapt to changes in the data and learn the characteristics of input signal [8]. Also, ANNs consume less storage and computation than SVMs [9]. Mostly used classifiers based on ANN are MLP and RBFN. B.K. Verma [10] presented a system for HCR using MLP and RBFN networks in the task of handwritten Hindi character recognition. The error back propagation algorithm was used to train the MLP networks. J. Sutha et al. in [11] showed the effectiveness of MLP for Tamil HCR using the Fourier descriptor features. R. Gheroie et al. in [12] proposed handwritten Farsi character recognition using MLP trained with error back propagation algorithm.

Similar shaped characters are difficult to differentiate because of very minor variations in their structures. In [13] T. Wakabayashi et al. proposed an F-Ratio (Fisher Ratio) based feature extraction method to improve results of similar shaped characters. They considered pairs of similar shaped characters of different scripts like English, Arabic/Persian, Devnagri, etc. and used QDF for recognition purpose. QDF suffers from limitation of minimum required size of dataset. F. Yang et al. in [14] proposed a method that combines both structural and statistical features of characters for similar handwritten Chinese character recognition.

As it can be seen that various feature extraction methods and classifiers have been used for character recognition by researchers that are suitable for their work, we propose a novel feature set that is expected to perform well for this application. In this paper, the features are extracted on the basis of character geometry, which are then fed to each of the selected ML algorithms for recognition of SSHHC.

3. MACHINE LEARNING CONCEPTS

Machine learning [15] is a scientific discipline that deals with the design and development of algorithms that allow computers to develop behaviours based on empirical data. ML algorithms, in this application, are used to map the instances of the handwritten character samples to predefined classes.

3.1. Machine Learning Algorithms

For this application of SSHHC recognition, we use the below mentioned ML algorithms that have been implemented in WEKA 3.7.0[16]: WEKA (Waikato Environment for Knowledge Analysis) is JAVA based open source simulator. These algorithms have been found performing very well for most of the applications and have been widely used by researchers. Brief description of these algorithms is as follows:

3.1.1. Bayesian Network

A Bayesian Network [17] or a Belief Network is a probabilistic model in the form of directed acyclic graphs (DAG) that represents a set of random variables by its nodes and their correlations by its edges. Bayesian Networks has an advantage that they visually represent all the relationships between the variables in the system via connecting arcs. Also, they can handle situations where the data set is incomplete.

3.1.2. Radial Basis Function Network

An RBFN [18] is an artificial neural network which uses radial basis functions as activation functions. Due to its non-linear approximation properties, RBF Networks are able to model complex mapping. RBF Networks do not suffer from the issues of local minima because the parameters required to be adjusted are the linear mappings from hidden layer to output layer.

3.1.3. Multilayer Perceptron

An MLP [19] is a feed forward artificial neural network that computes a single output from multiple real-valued inputs by forming a linear combination according to its input weights and then putting the output through some nonlinear activation function (mainly Sigmoid). MLP is a universal function approximator and is highly efficient for solving complex problems due to the presence of more than one hidden layer.

3.1.4. C4.5

C4.5 [20] is an extension of Ross Quinlan's earlier ID3 algorithm. It builds decision trees from a set of training data using the concept of information gain and entropy. C4.5 uses a white box model due to which the explanation of results is easy to understand. Also, it performs well with even large amount of data.

3.2. Feature Reduction

Feature extraction is the task to detect and isolate various desired attributes (features) of an object in an image, which maximizes the recognition rate with the least amount of elements. However, training the classifier with maximum number of features obtained is not always the best option, as the irrelevant or redundant features can cause negative impact on a classifier's performance [21] and at the same time, the build classifier can be computationally complex.

Feature reduction or feature selection is the technique of selecting a subset of relevant features to improve the performance of learning models by speeding up the learning process and reducing computational complexities. Two feature reduction methods that have been chosen for this application are CFS [22] and CON [23] as these methods have been widely used by researchers for feature reduction.

4. EXPERIMENTAL METHODOLOGY

The following sections describe the data-set, pre-processing and feature extraction adopted in our proposed work of recognition of SSHHC.

4.1. Dataset Creation

Dataset is created by asking the candidates of different age groups to write the similar shaped characters (□, □; □, □; □, □) several times in their handwriting on white plain sheets. These image samples are scanned using HP Scanjet G2410 with resolution of 1200 x 1200 dpi. Each character is cropped and stored in .jpg format using MS Paint. Thus this dataset consists of isolated handwritten Hindi characters that are to be recognized using ML algorithms. Using these character samples, three datasets are created as described below:

Dataset 1 consists of only 100 samples of the target pair (□, □).

Dataset 2 consists of increased samples of the same target pair, i.e. size of training dataset is increased to 342 samples by adding more samples of the target class (from other persons) in the training dataset. More samples are added in order to analyze the impact of increase in number of samples on the relative performance of ML algorithms.

Dataset 3 consists of samples of both the target and non-target class, i.e. other similar shaped character pairs (like □, □; □, □) are also added to the dataset (making 500 samples in total) with which the ML algorithms are trained. Non-target class characters are added to test the ability of ML classifiers for target characters among different characters. A few samples of the entire dataset are shown in Figure 1.

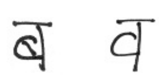

Figure. 1 Samples of Handwritten Hindi Characters

4.2. Performance Metrics

Performance of the classifiers is evaluated on the basis of the metrics described below:

i Precision: Proportion of the examples which truly have class x among all those which were classified as class x.

ii Misclassification Rate: Number of instances that were classified incorrectly out of the total instances.

iii Model Build Time: Time taken to train a classifier on a given data set.

4.3. Pre-processing

Following pre-processing steps are applied to the scanned character images:

i First each RGB character image, after converting to gray scale, is binarized through thresholding.

ii The image is inverted such that the background is black and foreground is white.

iii Then shortest matrix that fits the entire character skeleton for each image is obtained and this is termed as universe of discourse.

iv Finally, the spurious pixels are removed from the image followed by skeletonization.

4.4. Feature Extraction

After pre-processing, features for each character image are extracted based on the character geometry using the technique described in [24]. The features are based on the basic line types that form the skeleton of the character. Each pixel in the image is traversed. Individual line segments, their directions and intersection points are identified from an isolated character image.

For this, the image matrix is initially divided into nine zones and the number, length and type of lines and intersections present in each zone are determined. The line type can be: Horizontal, Vertical, Right Diagonal and Left Diagonal. For each zone following features are extracted. It results into a feature vector of length 9 for each zone:

i. Number of horizontal lines

ii. Number of vertical lines

iii. Number of Right diagonal lines

iv. Number of Left diagonal lines

v. Normalized Length of all horizontal lines

vi. Normalized Length of all vertical lines

vii. Normalized Length of all right diagonal lines

viii. Normalized Length of all left diagonal lines

ix. Number of intersection points.

A total of 81 (9x9) zonal features are obtained. After zonal feature extraction, four additional features are extracted for the entire image based on the regional properties namely:

i. Euler Number: It is defined as the difference of Number of Objects and Number of holes in the image

ii. Eccentricity: It is defined as the ratio of the distance between the centre of the ellipse and its major axis length

iii. Orientation: It is the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the region

iv. Extent: It is defined as the ratio of pixels in the region to the pixels in the total bounding box

Pre-processing and feature extraction has been implemented in MATLAB 7.9.0 as described in [25]. For each character image, 85 features are obtained in total.

5. RESULTS AND DISCUSSION

Based on the above described feature extraction technique, experiments are conducted for one similar shaped handwritten Hindi character pair ((□, □): target pair). System configuration on which the data set creation, feature extraction, training and testing is done has AMD dual core processor at 2.10 GHz, 3 GB RAM and Windows 7 Enterprise (32-bit) OS. Samples are replicated for the proper training of the classifiers. The complete feature set, consisting of 85 features, created on the basis of geometry of the character is given to each of the ML algorithms for their training. Testing is done on a file consisting of 72 samples of the target pair and same features as of training file are used for each test experiment.

Table 1 shows the performance analysis of the ML algorithms when tested on 72 samples of testing file after the algorithms are trained with Dataset 1, 2 and 3. Precision and Misclassification Rate are taken in percentage and Model Build Time in seconds.

Table 1. Impact of increase in number of samples of the target class and non-target class in the training dataset using complete feature set

ML Algorithms	**Dataset 1**			**Dataset 2**			**Dataset 3**		
	Precision	**Mis-classification Rate**	**Model build Time**	**Precision**	**Mis-classification Rate**	**Model build Time**	**Precision**	**Mis-classification Rate**	**Model build Time**
Bayesian Network	93.9	7.31	0.03	96.6	3.65	0.02	100	0	0.03
RBFN	85.5	14.63	0.12	83.2	17.07	0.1	91	12.19	3.25
MLP	93.6	7.31	10.03	100	0	35.58	98.8	2.43	51.58
C4.5	91.1	9.75	0.03	98.8	1.21	0.07	96.3	4.87	0.1

When the number of samples is increased for the target character pair and classifiers are trained with Dataset 2, performance of C4.5, MLP and Bayesian Network improves as obvious from Table 1. From these results, we expect that performance of these ML algorithms will improve further if the size of training dataset is increased.

A practical ML classifier has to come across a number of different characters, other than the target characters. From Table 1, it can be seen that when trained with Dataset 3 of samples of the target and non-target class characters, Bayesian Network gives better performance as compared to all other algorithms. This shows that Bayesian Network gives consistent performance and we do not perceive any impact on the performance of Bayesian Network classifier when different character samples are added in the training dataset.

It can also be noticed that Bayesian Network and C4.5 build quickly when trained with each of the training file. Build time for RBFN is acceptable; however MLP builds slowly which means that size of the training dataset may inversely affect the speed of MLP for its training. Thus, Bayesian Network and C4.5 can be trained quickly even with a large training set.

Further, to reduce the dimensionality of the feature vector, we apply CFS and CON feature reduction techniques using the Best first search method, as it is a commonly used method and yields good results. Table 2 shows the list of 11 features obtained using CFS. With this feature set, experiments are conducted with Dataset 1, 2 and 3 and the performance of the chosen ML algorithms is analyzed when tested on the same test file with 72 samples but with the selected 11 features.

Table 2. List of CFS Features

Feature Number	Feature Name
6	Normalized Length of all vertical lines in Zone1
10	Number of horizontal lines in Zone2
18	Number of Intersection points in Zone 2
33	Normalized Length of all vertical lines in Zone 4
35	Normalized Length of all left diagonal lines in Zone 4
36	Number of Intersection points in Zone 4
40	Number of left diagonal lines in Zone 5
60	Normalized Length of all vertical lines in Zone 7
71	Normalized Length of all left diagonal lines in Zone 8
84	Orientation
85	Extent

It is observed from Table 3 that performance of Bayesian Network consistently improved using CFS features. As Bayesian Network works on joint probability distribution [16], less number of features results in less number of parent nodes required for probability computations. Also, reduction in feature set has led to reduced model build time, which is almost half in case of Bayesian Network and C4.5, while MLP's build time has significantly reduced.

Table 3. Impact of Correlation-based Feature Reduction on the performance of ML Algorithms

ML Algorithms	100 Samples			342 Samples			500 Samples		
	Precision	Mis-classification Rate	Model build Time	Precision	Mis-classification Rate	Model build Time	Precision	Mis-classification Rate	Model build Time
Bayesian Network	91.7	8.54	0.01	96.4	3.66	0.02	98.8	2.44	0.01
RBFN	89.3	12.2	0.06	69.7	32.93	0.09	95.1	7.32	6.61
MLP	91.1	9.75	0.49	98.8	1.22	1.51	93.8	8.54	2.3
C4.5	91.1	9.75	0.04	96.4	3.66	0.02	96.3	4.87	0.02

Then, the original feature set is subjected to CON feature reduction method and 4 features are obtained as mentioned in Table 4.

Table 4. List of CON Features

Feature Number	Feature Name
1	Number of Horizontal lines in Zone 1
36	Number of Intersection points in Zone 4
83	Eccentricity
84	Orientation

Table 5 details the performance of ML algorithms when trained with Dataset 1, 2 and 3 using CON features. It is observed that C4.5 performed exceptionally well with CON feature set. As the number of features is fairly reduced, entropy is reduced and information gain [19] is maximized, thereby enhancing the classifier's ability to make correct decisions. With such small number of features, Bayesian Network builds immediately when trained, in no time. C4.5 is also found be very quick in training.

Table 5. Impact of Consistency-based Feature Reduction on the performance of ML Algorithms

ML Algorithms	100 Samples			342 Samples			500 Samples		
	Preci sion	Mis-classifi cation Rate	Model build Time	Preci sion	Mis-classific ation Rate	Model build Time	Precis ion	Mis-classific ation Rate	Model build Time
Bayesian Network	86.8	15.85	0	84.2	17.07	0	90.9	18.29	0
RBFN	84.7	15.85	0.04	83.3	18.29	0.06	90.9	20.73	0.78
MLP	87.6	14.63	0.28	88.6	12.19	0.92	85.2	40.24	1.18
C4.5	89.3	12.19	0.01	98.8	1.22	0.02	100	3.66	0.02

From the above experiments, we can see that overall Bayesian Network and C4.5 have given better performance as compared to the other two ML algorithms. So, in order to analyze the overall error rate, we compare Bayesian Network and C4.5 using Dataset 3 on the basis of misclassification rate.

Based on Full, CFS and CON feature set, misclassification rate of Bayesian Network and C4.5 is shown in Figure 2. It is clearly evident that if misclassification rate is of high priority, then Bayesian Network with CFS features can be a better option as its performance consistently improved even when non-target characters are present in the target class, but at the stake of computational complexity as number of features is more. However, if computational complexity is to be taken into account, then C4.5 with CON can be a good choice, but at the expense of misclassification rate.

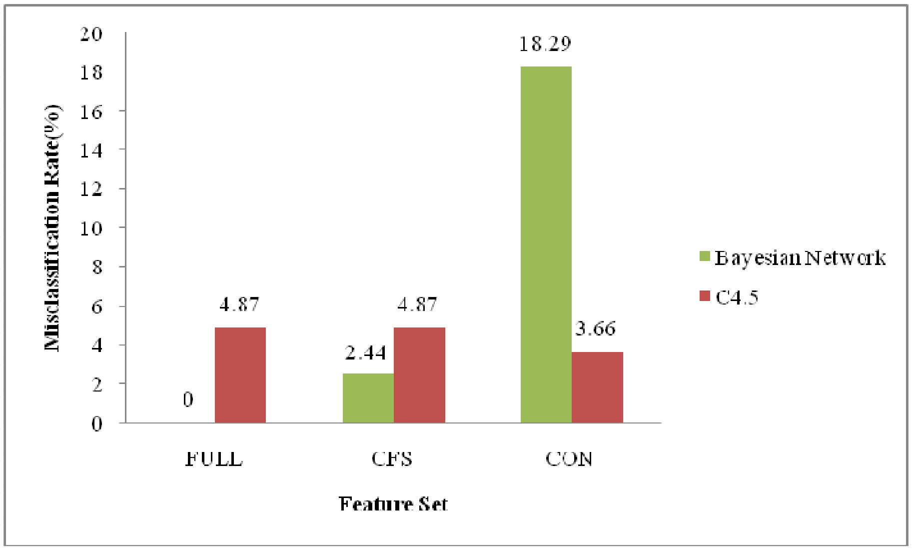

Figure 2. Misclassification Rate of Bayesian Network and C4.5 with FULL, CFS and CON Features Sets

5. CONCLUSION AND FUTURE WORK

In character recognition, misclassification occurs more for similar shaped handwritten characters. Traditional techniques used for character recognition bear certain limitations [4], [5], [6]. In this paper, the recognition of Similar Shaped Handwritten Hindi Characters (SSHHC) is presented using four ML algorithms namely Bayesian Network, RBFN, MLP and C4.5 and their performance is analyzed and compared. It is found that MLP gives better performance with increase in number of samples of the same target pair in the training dataset. Bayesian Network shows better performance as compared to all other ML algorithms even when it is trained with dataset consisting of non-target characters present along with the target characters.

A total of 85 features are extracted based on character geometry. So, CFS and CON feature reduction techniques are applied to reduce this high dimensionality of the feature vector. It is found out that Bayesian Network with CON features is a better choice if misclassification rate is of more significance though with higher computational complexity. C4.5 is to be selected if classifier needs to be computationally efficient, as it is using less number of features. Also, Bayesian Network, due to its significantly less model build time, can be retrained during classification and thus can be applied to online classification problems. This work can further be extended by taking samples of all Hindi characters and devising new techniques for feature extraction, those are highly correlated with target class of characters.

REFERENCES

[1] T.M. Mitchell, "*Machine Learning*", McGraw-Hill Education (ISE Editions), December 1997.

[2] M. Zafar, D. Mohamad, M.M. Anwar, "Recognition of Online Isolated Handwritten Characters by Backpropagation Neural Nets Using Sub-Character Primitive Features", IEEE Multitopic Conference(INMIC), 2006, pp. 157 - 162

[3] G. Vamvakas , B. Gatos, S. Petridis, N. Stamatopoulos , "An Efficient Feature Extraction and Dimensionality Reduction Scheme for Isolated Greek Handwritten Character Recognition", IEEE Ninth International Conference on Document Analysis and Recognition(ICDAR), 2007, vol. 2, pp. 1073 – 1077.

[4] J.R. Prasad, U.V. Kulkarni, R.S. Prasad, “Offline handwritten character recognition of Gujrati script using pattern matching”, IEEE 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication, 2009, pp. 611-615.

[5] H.S. Park, S.W. Lee, “An HMMRF-Based Statistical Approach for Off-line Handwritten Character Recognition”, IEEE Proceedings of the 13th International Conference on Pattern Recognition, 1996, vol. 2, pp. 320 – 324.

[6] C.L. Liu, H. Sako, H. Fujisawa, “Performance evaluation of pattern classifiers for handwritten character recognition”, International Journal on Document Analysis and Recognition (IJDAR), 2002, vol. 4, pp. 191–204.

[7] N. Araki, M. Okuzaki, Y. Konishi , H. Ishigaki , “A Statistical Approach for Handwritten Character Recognition Using Bayesian Filter”, IEEE 3rd International Conference on Innovative Computing Information and Control, 2008, pp. 194 – 194.

[8] N. Arica, F.T. Yarman-Vural, “An Overview Of Character Recognition Focused On Off-line Handwriting”, IEEE Transactions on Systems, Man, and Cybernetics, 2001, vol. 31, pp. 216 – 233.

[9] F. Kahraman, A. Capar, A. Ayvaci, H. Demirel, M. Gokmen, “Comparison of SVM and ANN performance for handwritten character classification”, Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference, 2004, pp. 615 – 618.

[10] B.K. Verma, “Handwritten Hindi Character recognition Using Multilayer Perceptron and Radial Basis Function Neural Networks,” Proceedings of IEEE International conference on Neural Networks, 1995, vol. 4, pp. 2111-2115.

[11] J. Sutha, N. Ramaraj, “Neural Network Based Offline Tamil Handwritten Character Recognition System”, IEEE International Conference on Computational Intelligence and Multimedia Applications, 2007, vol. 2, pp. 446 – 450.

[12] R. Gharoie, M. Farajpoor, “Handwritten Farsi Character Recognition using Artificial Neural Network”, International Journal of Computer Science and Information Security, 2009, vol. 4.

[13] T. Wakabayashi, U. Pal, F. Kimura, Y. Miyake, “F-ratio Based Weighted Feature Extraction for Similar Shape Character Recognition” IEEE 10th International Conference on Document Analysis and Recognition (ICDAR), 2009, pp. 196-200.

[14] F. Yang, X.D. Tian, X. Zhang, X.B. Jia, “An Improved Method for Similar Handwritten Chinese Character Recognition”, IEEE Third International Symposium on Intelligent Information Technology and Security Informatics (IITSI), 2010, pp. 419 – 422.

[15] N.J. Nilsson, “Introduction to Machine learning”, An early draft of a proposed textbook, Robotics Lab, Deptt of Computer Science, Stanford University, Stanford, CA 94305, 1998.

[16] WEKA 3: Data Mining With Open Source Machine Learning Software in JAVA

URL: http://www.cs.waikato.ac.nz/ml/weka/.

[17] N. Friedman, D. Geiger, M. Goldszmidt, “Bayesian Network Classifiers, Machine learning, 1997, pp. 131-163.

[18] A. G. Bors, "Introduction of the Radial Basis Function (RBF) Networks," Online Symposium for Electronics Engineers, 2001, vol. 1, pp. 1-7.

[19] L. Noriega, “Multilayer Perceptron Tutorial”, School of Computing, Staffordshire University, UK, 2005, pp. 1-12.

[20] S. Drazin, M. Montag, “Decision Tree analysis using Weka”, Machine Learning- Project II, University of Miami, pp. 1-3.

[21] N. Williams, S. Zander, G. Armitage, “A Preliminary Performance Comparison of Five Machine Learning Algorithms for Practical IP Traffic Flow Classification”, ACM SIGCOMM Computer Communication Review, 2006, vol. 36, pp. 7-15.

[22] M. Dash, H.Liu, “Consistency-based Search in Feature Selection”, Artificial intelligence, 2003, vol. 151, pp. 155-176.

[23] M. Hall, “Correlation-based Feature Selection for Machine Learning”, PHD Thesis, Deptt of Computer Science, Waikato University, Hamilton, New Zealand, 1998.

[24] M. Blumenstein, B. Verma, H. Basli, “A Novel Feature Extraction Technique for the Recognition of Segmented Handwritten Characters”, IEEE Proceedings of Seventh International Conference on Document Analysis and Recognition, 2003, vol. 1, pp. 137 – 141.

[25] MATLAB Central

URL: http://www.mathworks.com/matlabcentral/fileexchange/24624-feature-extraction-for-character-recognition