

Prediction of Writer Using Tamil Handwritten Document Image Based on Pooled Features

T. Thendral, M. S. Vijaya, S. Karpagavalli

Abstract—Tamil handwritten document is taken as a key source of data to identify the writer. Tamil is a classical language which has 247 characters include compound characters, consonants, vowels and special character. Most characters of Tamil are multifaceted in nature. Handwriting is a unique feature of an individual. Writer may change their handwritings according to their frame of mind and this place a risky challenge in identifying the writer. A new discriminative model with pooled features of handwriting is proposed and implemented using support vector machine. It has been reported on 100% of prediction accuracy by RBF and polynomial kernel based classification model.

Keywords—Classification, Feature extraction, Support vector machine, Training, Writer.

I. INTRODUCTION

WRITERS' distinctiveness can be authenticated based on physiological characteristics and behavioral characteristics. The physiological characteristics like fingerprints, face, iris, retina, hand geometry and the behavioral characteristics like a voice, signature, gait handwriting. Handwriting can exhibit in the behavioral characteristics of an individual. The significance and scope of writer identification is becoming more prominent in these days. Writer identification is of a primordial importance of forensic document examination as it helps experts on deliberating on the authenticity of a certain document. Signature verification is a specific case of writer identification system, which can be used for automatic personal identification and to detect forgeries as signature is judicially accepted proof of identity of an individual in legal and commercial environments, document analysis and access control.

Writer identification can be administered in two behaviors: text-dependent and text-independent. In text-dependent method, writers have to write the identical text to perform identification. In text independent method different text may be used to establish the identity of writers [1]. The text capturing modes can be normally classified into two types as online and offline. In online, the writing behavior is directly captured from the writer and converted to a sequence of signals using a transducer device but in offline mode, the

handwritten text in the form of scanned images is used for identification process. On-line writer identification is extensively considered as more challenging than off-line; because, it contains more information about the writing style of a person, such as pressure, speed, angle which is not available in the off-line mode.

For the past three decades, enormous effort has been made on handwritten recognition and writer identification to provide a solution for processing large volumes of data automatically in a large variety of scientific and business applications [2]. Recent advances in computational engineering, artificial intelligence, data mining, image processing, pattern recognition and machine learning have proved that it is possible to automate writer identification. Writer identification has been studied extensively for several languages like Chinese, English, Arabic, Persian, Kurdi, Jawi and Urdu [3] and also for several Indian languages like Hindi, Oriya, Bangla, Kannada, and Malayalam. Various approaches like a pattern matching, statistical, machine learning have been adopted for processing the documents and developing the writer identification models for these languages.

Text writes in Tamil is the key data onto this research. Tamil letters have 12 vowels called soul-letters (உயிரெழுத்து uyireluttu), 18 consonants called body-letters (மெய்யெழுத்து meyyeluttu) and one character called aytam (ஆய்தம்). The complete Tamil letters has thirty one independent forms and additional 216 combinant letters combined to have a total 247 combinations (உயிர்மெய்யெழுத்து uyirmeyyeluttu) of a consonant and a vowel, a mute consonant, or a vowel alone. Some Tamil letters namely க்ஷ ksa and ஸ்ரீ sri which is borrowed from Sanskrit language. Accordingly, Tamil handwriting identification has more complexities than any other language.

The difficulties involved in Tamil alphabet system need to be analyzed and to be taken into account while modeling the writer identification task. These are some of the issues that need to be addressed during identification of individuals based on their handwriting. In our previous work, limited words are used to predict the writer [4], [5]. The text with limited words will not be sufficient to predict the writer accurately. So the text with more sentences is required for training the classifier.

The above challenges can be managed by considering the appropriate features of the handwriting. Discovering of Tamil writer has been demonstrated and implemented using a machine learning technique. Support vector machines are used to train the model and its enactment is appraised to predict the writer.

T. Thendral is with the PSGR Krishnammal College for Women, Peelamedu, Coimbatore-641004 (phone: +91 9842946043; e-mail: thendral.14@gmail.com).

M. S. Vijaya and S. Karpagavalli are with the Computer Science (PG) Department, PSGR Krishnammal College for Women, Peelamedu, Coimbatore-641004 (e-mail: msvijaya@psgrkc.com, karpagavalli@psgrkc.com).

II. PROPOSED WRITER IDENTIFICATION MODEL

Individual exclusivity can be substantiated using a common technique namely handwriting recognition. The fundamental property of handwriting is that there exists difference in individual's writing which make writer identification possible. The writer's invariants, reflecting the writing style or writing individuality of his/her handwriting is defined as the set of similar patterns extracted from his/her handwriting [6].

The key center of this exploration is to deliver a resourceful machine learning based interpretation of the problem of writer identification with adapting to pattern classification problems. For exhibiting instinctive writer identification, the important tasks such as corpus preparation, preprocessing, feature extraction, training the model and prediction of writer based on his/her handwriting using the learned model have been carried out.

A. Corpus Preparation

The research work proposed here for modeling writer identification is based on offline text dependent approaches. A set of 100 identical tamil documents write by 300 different writers have been collected and the handwritten documents are scanned using a scanner of resolution 300 dpi. A total of 30000 JPEG paragraphs images are separated from each document image and a corpus is prepared. The writers of different age groups are considered here. The individuals have been identified based on various factors like education level, age, gender and locality. Sample of text dependent documents are shown in Fig. 1

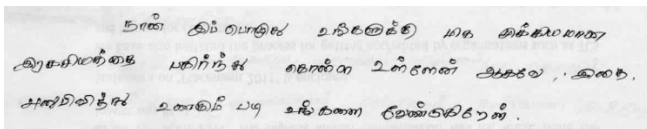


Fig. 1 (a) Text Dependent – Sample Data Writer1

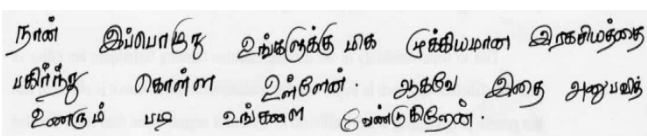


Fig. 1 (b) Text Dependent – Sample Data Writer2

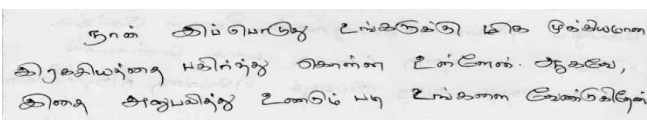


Fig. 1 (c) Text Dependent – Sample Data Writer3

B. Preprocessing

Preprocessing is a significant task in data mining technique that encompasses transmuting inexplicable data onto an explainable format. It was observed that preprocessing and features describing the pattern of handwriting are two important aspects in writer identification, which in turn play a major role in improving the accuracy of the prediction model. Normalisation is a technique used here to preprocess the data [7].

1. Normalisation of Handwriting Image

Normalisation requires two different stages. First the detection and correction of the skewed words with the handwriting images has been performed [7]. Then, the space between horizontal and vertical lines of a text has been normalised.

2. Skewed Words Normalisation

Handwriting images are not assured to be exact during scanning as it is in the document. The disposition of different characters, words and lines in handwriting images may distress the writer identification. To overcome this skew-normalisation procedure is implemented. This procedure can be performed using the following steps,

- Step1. In a handwriting image horizontal projection profile (HPP) method is applied to detect text lines and empty spaces.
- Step2. Closing procedure is applied to the image using a 3_3 structuring elements. (The middle row of the element is set so as to close the image in the horizontal direction to avoid joining text lines is called closing procedure).
- Step3. Select the connected components.
- Step4. In that the minimum, maximum and mean connected component heights is calculated.
- Step5. Filter out the smallest 5% of height in the text to eliminate punctuation, quotes and so on.
- Step6. Formerly eliminate components with a height > 2 _mean height to remove components which are even now connected with more than one text line.
- Step7. Then perform the following, in the remaining connected component.
- Step8. Copy the component into a blank image, in which the image has the component bounding box size.
- Step9. In that the line fit on the connected component is performed.
- Step10. Finally, base line fitting in which the base lines are computed from the HPP of the deskewed image.

In conclusion execute base line fitting to the deskewed handwriting image, to produce horizontal text lines.

C. Computation of Features

Feature extraction is another important task in modeling writer identification [8]. The features describing the writing pattern of an individual are extracted from the preprocessed images. The distinct features such as Gabor Filter, GLCM (Gray Level Co-occurrence Matrix), GGD (Generalized Gaussian Density), Contourlet GGD, Directional features are computed from the document images. These features are used to prepare the training dataset [9].

1. Gabor-Filter

Gabor-filter is a feature specifically used for machine visualization applications such as handwritten identification, edge detection, object detection and so on [10]. This filter is applied for an image of setting different frequencies and orientations which may helpful for extracting useful features. The frequency applied here is 4, 8, 16 and 32 cycles/degree.

The two-dimensional Gabor function is applied to detect the features of an image.

2. Gray Level Cooccurrence Matrix (GLCM)

G is an image where number of grey levels is equal to the number of rows and columns of a matrix in GLCM [11]. It is de-fined over an image to be the distribution of co-occurring values of the given offset. It is a way of extracting second order statistical features. It is used to measure the spatial relationships between pixels. This method is based on the belief that texture information is contained in such relationships. The GLCMs are constructed by mapping the grey level co-occurrence probabilities based on spatial relations of pixels in different angular direction. Greycomatrix function creates the GLCM by calculating how often pixels with grey-level value i occurs horizontally adjacent to a pixels with the value j. Each element (i, j) in GLCM specifies the number of times that the pixels with values I occurred horizontally adjacent to a pixels with value j.

The grey co matrix can be calculated based on the scaled version of the image. If an image is a binary image then greycomatrix function scales the images of two grey-levels. If images are an intensity images, greycomatrix function scales the image to eight grey-level. It can specify the number of Grey level greycomatrix functions to scale the image of using the 'NumLevel' parameter, and the grey co matrix scales the values using the 'GreyLimits' parameter.

Element (1, 1) in the GLCM contains the values 1 because there is only one in-stances in the image where two, horizontally adjacent pixels have the values 1 and 1. Element (1, 2) in the GLCM contains the value 2 because there are two instances in the images where two, horizontally adjacent pixels have the values 1 and 2. Greycomatrix functions to continue this process to fill in all the values of the GLCM.

Using this GLCM, 22 texture features are extracted and are illustrated below.

Energy

Textures disorders can be detected using this approach. Energy reaches a maximum value equal to one.

$$\text{energy (ene)} = \sum_i \sum_j g_{ij}^2 \quad (1)$$

Entropy

The image is not texturally uniform then the value of entropy will be high and many GLCM elements have very small values. Complex textures show very high value of entropy.

$$\text{entropy (ent)} = - \sum_i \sum_j g_{ij} \log_2 g_{ij} \quad (2)$$

Contrast

The spatial frequency of an image and difference moment of GLCM can be measured using this approach. The difference between the highest and the lowest values of a contiguous set of pixels is said to be contrast. The amount of

local variation present in the image can be measured using contrast.

$$\text{contrast (con)} = \sum_i \sum_j (i - j)^2 g_{ij} \quad (3)$$

Variance

Variance increases when the gray level values differ from their mean [12].

$$\text{variance (var)} = \sum_i \sum_j (i - \mu)^2 g_{ij} \quad \text{where } \mu \text{ is the mean of } g_{ij} \quad (4)$$

Homogeneity

Homogeneity weights values is inverse to the contrast weight, with weights decreasing exponentially away from the diagonal.

$$\text{homogeneity (hom)} = \sum_i \sum_j \frac{1}{1+(i-j)^2} g_{ij} \quad (5)$$

Correlation

GLCM correlation has different calculation from the other texture measures in which 0 is uncorrelated, 1 is perfectly correlated.

$$\text{correlation (cor)} = \frac{\sum_i \sum_j (ij) g_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y} g_{ij} \quad (6)$$

Autocorrelation

The coarseness can be measured using autocorrelation function. It is used to evaluate the linear spatial relationships between primitives. The autocorrelation increases and decreases periodically with distance according to the primitives.

Sum Average:

$$\text{sum average (sa)} = \sum_{i=2}^{2N_g} i g_{x+y}(i) \quad (7)$$

Sum Entropy

$$\text{sum entropy (se)} = - \sum_{i=2}^{2N_g} i g_{x+y}(i) \log\{g_{x+y}(i)\} \quad (8)$$

Sum Variance

$$\text{sum variance (sv)} = \sum_{i=2}^{2N_g} (i - \text{sa})^2 g_{x+y}(i) \quad (9)$$

Difference Variance

$$\text{difference variance} = \text{variance of } g_{x-y} \quad (10)$$

Difference Entropy

$$\text{difference entropy (se)} = - \sum_{i=0}^{N_g-1} g_{x-y}(i) \log\{g_{x-y}(i)\} \quad (11)$$

Information Measures of Correlation

i) Information Measures of Correlation 1 (IMC1)

$$\text{IMC1} = \frac{\text{HXY} - \text{HXY1}}{\max\{\text{HX}, \text{HY}\}} \quad (12)$$

ii) Information Measures of Correlation 2 (IMC2)

$$\text{IMC2} = \sqrt{(1 - \exp[-2.0(\text{HXY2} - \text{HXY})])} \quad (13)$$

where,

$$HX = -\sum_i \sum_j g_{ij} \log_2 g_{ij} \text{ where } HX \text{ \& } HY \text{ are entropies of } g_x \text{ and } g_y \text{ (14)}$$

$$HXY1 = -\sum_i \sum_j g_{ij} \log_2 \{g_x(i)g_y(j)\} \text{ (15)}$$

Cluster Shade

$$\text{Shade} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i + j - \mu_x - \mu_y\}^3 * P(i, j) \text{ (16)}$$

Cluster Prominence

$$\text{Prom} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i + j - \mu_x - \mu_y\}^4 * P(i, j) \text{ (17)}$$

A GLCM is a matrix where the number of rows and columns is equal to the number of gray level G in the image, where

$$\mu_x = \sum_{i=0}^{G-1} iP_x(i) \\ \mu_y = \sum_{j=0}^{G-1} jP_y(j) \text{ (18)}$$

dissimilarity

$$\text{diss} = \sum P_{ij} * |i - j| \text{ (19)}$$

3. Generalized Gaussian Density

In a handwriting image Wavelet-based GGD method is used to establish corresponding wavelet-based GGD model. Wavelets here refers to mathematical functions which have parameters α, β are regarded as the features of the handwritten image [13]. Using different frequency components the GGD models cut the image of regions for each handwritten image. Every region is called wavelet subband. For each wavelet subband the probability is maximized for the estimated parameters α, β which are optimal for improving accuracy of writer identification. The probability is estimated as,

$$P(\{\alpha, \beta\})/X \text{ (20)}$$

The coefficients of wavelet subband is $X = \{x_1, x_2, \dots\}$. After the error probability the function is defined as,

$$L(X|\{\alpha, \beta\}) = \log \prod_{i=1}^N P(x_i|\{\alpha, \beta\}) \text{ (21)}$$

4. Contourlet Generalized Gaussian Density

In GGD, wavelet based GGD model is used to transform the image into subbands in different frequency and orientation. This wavelet can capture only limited directional operation is a big disadvantage of image analysis. To overcome this issue and to capture the geometrical structures such as edges or contours of the handwritten image can be done by using multiscale and directional representations of an image. Contourlet generalized Gaussian [14] can deal effectively with images having smooth contours be an efficient two-dimensional multiscale and directional filter bank. The double filter bank called pyramidal directional filter banks (PDFB) is used to implement the contourlet. In PDFB, the 9-7 filter bank is applied to decompose the images. After that directional filter bank is applied to analyze the image of a four directional sub bands. The GGD model and its parameters are used to

represent the contourlet subband. To estimate the parameters α, β by converting the subband images into multi-dimensional vector and it is defined as,

$$L(x; \alpha, \beta) = \log \prod_{i=1}^L p(x_i; \alpha, \beta) \text{ (22)}$$

5. Directional Feature

The normalized feature vector for classification is obtained using the directional features. Here first the input image is characterized as four types, such as vertical line, horizontal line, left diagonal and right diagonal [15]. The values are calculated from the four directions, as fraction of the distance traversed across the image. If the transition is computed from left to right, a transition found close to the left is assigned a high value compared to a transition computed further to the right. A maximum value (MAX) is defined as the largest number of transitions that is recorded in each direction. The remaining MAX transitions are assigned values of 0.

For a particular direction the transition value is calculated. To compute the directional transition, the transition value is divided between a predetermined number which is assigned as 10. For one transition eight features can be obtained. Then this process is repeated and it can obtain 256 features, using the following formula,

$$\text{nrFeatures} * \text{nrTransitions} * \text{nrVectors} * \text{resampledMatrixHeight} \\ \text{(Width)}$$

D. Training Dataset

The features extracted from each handwritten text image form a feature vector. The dimension of the feature vector here is 422. Since the corpus consists of 30000 paragraph images of ten writers, a set of 30000 feature vectors have been created. For each feature vector the class label is assigned from 1 to 10. Thus a training dataset with 30000 instances is developed for building writer identification models.

III. SUPPORT VECTOR MACHINE

Support Vector Machine is an intelligent tool that is being successfully applied to a wide range of pattern recognition problems. SVM is based on strong mathematical foundations and statistical learning theory but results in simple yet very powerful algorithms with high generalization power on unseen data. The SVM based writer identification models can be built using Mat lab/LIBSVM/SVMLight by tuning the various parameters like regularization parameters, gamma with different kernels. The standard SVM algorithm builds a binary classifier by constructing a hyper plane which separates out classes of data. SVM automatically identifies a subset of informative points called support vectors and uses them to represent the separating hyper plane.

The machine is presented with a set of training examples, (x_i, y_i) where the x_i is the real world data instances and the y_i are the labels indicating which classes the in-instance belongs to. For the two class pattern recognition problem, $y_i = +1$ or $y_i = -1$. A training example (x_i, y_i) is called positive if $y_i = +1$ and negative otherwise. SVMs construct a hyper plane that separates two classes and tries to achieve maximum separation

between the classes. Separating the classes of a large margin minimizes a bound on the expected generalization error.

When the number of classes is more than two, then the problem becomes multiclass classification. There are two approaches namely direct and indirect method for multiclass SVM. In indirect approaches, several binary SVM's are constructed and the classifier's output is combined with finding the final class [16]. In direct approaches, a single optimization formulation is considered for implementing multiclass SVM.

IV. EXPERIMENT AND RESULTS

The support vector machine based models is learned to use the normalized training dataset and the efficient classifier is built by tuning the SVM parameters. Subsequently the classifier can be used to predict the individual when his/her offline handwriting is given as input. Finally the writer identification system is developed by integrating the classifier in order to predict the writer of the document. The SVM training has been carried out using SVMlight [17] for multi classification with linear, polynomial and RBF kernels and with different settings for parameters such as C-the regularization parameter, d-degree of polynomial, and g-parameter gamma in RBF kernel [18].

The performance of trained models is evaluated using 10-fold cross validation. In that Number of iterations (I), Number of support vectors (SV), Number of correctly classified instances (CCI), Number of incorrectly classified instances (ICCI), predictive accuracy (PA) and learning time (LT) are observed. Prediction accuracy is the ratio of numbers of correctly classified instances and the total number of instances. Learning time is the time taken to build the model on the dataset. As far as writer identification is concerned predictive accuracy plays major roles than other concerns in predicting the identity of a writer based on handwriting. The performances of SVM classifiers are evaluated and compared.

The cross validation results in classification models based on SVM with linear kernels, SVM with RBF kernel and SVM with polynomial kernel are shown in Tables I-III. The average and comparative performance of classifiers is shown in Table IV.

TABLE I
SVM WITH LINEAR KERNEL

Linear SVM	C=1	C=10	C=26
No. of I	48	103	146
No. of SV	178	330	572
No. of CCI in %	24102	27656	28202
No. of ICCI in %	5898	2344	1798
Accuracy (%)	80.3%	92%	94%
Time Taken	4.57	4.76	5.10

TABLE II
SVM WITH RBF KERNEL

Parameter	C=0.5			C=1			C=5		
G	0.5	2	3	0.5	2	3	0.5	2	3
No. of I	3	3	3	3	3	3	3	3	3
No. of SV	152	198	153	162	188	139	192	167	146
No. of CCI in %	29997	29995	29999	29996	29998	29999	29995	30000	29999
No. of ICCI in %	3	5	1	4	2	1	5	0	1
Accuracy (%)	100	100	100	100	100	100	100	100	100
Time Taken	12.45	13.52	13.51	12.52	11.66	12.78	12.47	12.62	12.71

TABLE III
SVM WITH POLYNOMIAL KERNEL

Parameters	C=0.5		C=1		C=5	
D	1	2	1	2	1	2
No. of I	41	222	63	222	101	222
No. of SV	170	1780	206	1780	280	1780
No. of CCI in %	24003	29998	25812	29996	27052	29995
No. of ICCI in %	5997	2	4188	4	2948	5
Accuracy (%)	80	100	86	100	90.1	100
Time Taken	2536.49	8654.58	2838.27	8635.12	2624.16	7256.89

TABLE IV
AVERAGE PERFORMANCES OF THREE MODELS

Kernels	No. of I	No. of SV	No. of CCI	No. of ICCI	Accuracy	Learning time
Linear	146	572	28202	1798	94	5.10
RBF	3	146	29999	1	100	12.71
Polynomial	222	1780	29998	2	100	8654.58

The average and the comparative performance of the above SVM based prediction models is observed in terms of Number of iterations, Number of support vectors, Number of correctly classified instances, number of incorrectly classified instances, accuracy and learning time are illustrated in Figs. 2-4.

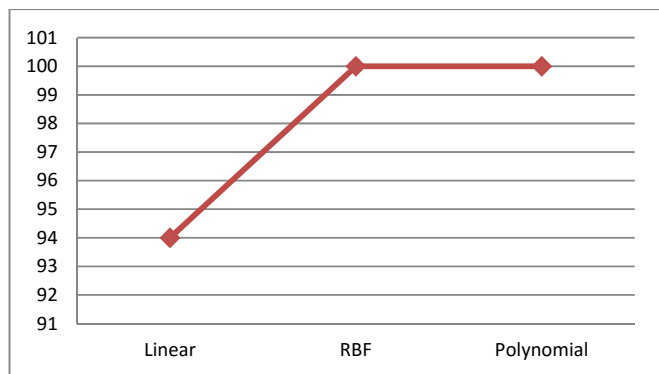


Fig. 2 Prediction Accuracy

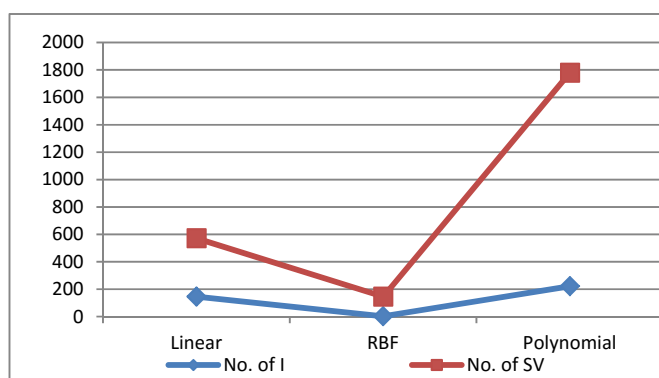


Fig. 3 No. of Iterations and Support Vectors

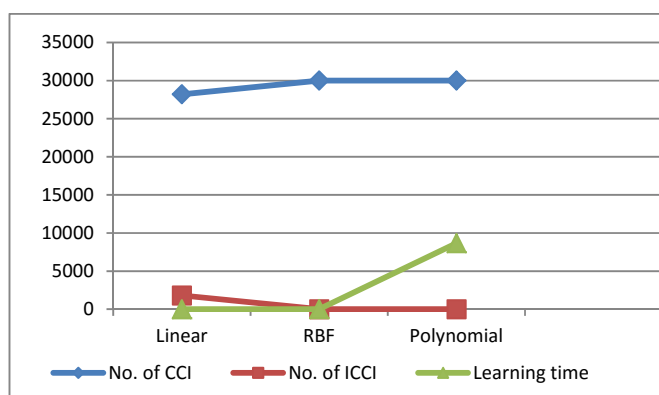


Fig. 4 Classified Instances and learning time

From the above qualified scrutiny it is observed that the RBF kernel and polynomial kernel based prediction model (100%) shows high accuracy and linear kernel (94%) SVMs and average time taken to construct the model is elevated in SVM with polynomial kernels (8654.58) than the other models. In this case of point of writer identification problems, the protrusive accuracy achieved its all-out in concluding the performance of the model. Hence it is concluded that SVM with RBF kernel and polynomial kernel (100%) based writer recognition model out performs well.

V. CONCLUSION

This research work proposes writer identification of tamil handwriting is achieved more effectively. Supervised learning technique namely support vector machine is used to build the model. Multiclassification task is applied to identify the writer. The implementation discovers SVM with RBF kernel and polynomial kernel based writer prediction model reaches its height of 100% accuracy. The proposed work achieved its maximum prominences while testing the data. In future, this research work continues to have its more achievements.

ACKNOWLEDGMENT

The authors are grateful to the UGC, New Delhi for having provided the fund to carry out this research work.

REFERENCES

- [1] Bulacu M and Schomaker L, "Text-independent writer identification and verification using textural and allographic features," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.701-717, April 2007.
- [2] Jagadeesh Kannan R, Prabhakar R, "A Comparative study of Optical Character Recognition for Tamil Script", European Journal of Scientific Research, Vol.35, No.4(2009), pp.570-582.
- [3] Wong Yee Leng & Siti Mariyam Shamsuddin, "Writer Identification for Chinese Handwriting", International Journal of Advances in Soft Computing and Its Applications, Vol. 2, No. 2, July 2010
- [4] Thendral T, Vijaya MS, Karpagavalli S, "Analysis of Tamil Character Writings and Identification of Writer Using Support Vector Machine", IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).
- [5] Thendral T, Vijaya MS, Karpagavalli S, "Supervised Learning Approach for Tamil Writer Identity Prediction Using Global and Local Features, International Journal of Research in Engineering and Technology (IJRET) Vol. 2, No. 5, 2013 ISSN 2277 - 4378, Page 204 - 208.
- [6] Sargur N. Srihari Ph.D., Sung-Hyuk Cha Ph.D., Hina Arora M.E. and Sangjik Lee M.S., "Individuality of Handwriting", J Forensic Sci, July 2002, Vol. 47, No. 4 Paper ID JFS2001227_474.
- [7] H. E. S. Said, G. S. Peake, T. N. Tan and K. D. Baker, "Writer Identification from Non-uniformly Skewed Handwriting Images", British Machine Vision Conference.
- [8] M. Blumenstein and X. Y. Liu, B. Verma, "A modified direction feature for cursive character recognition", IEEE International Joint Conference, Vol. 4, 25-29 July 2004, Page: 2983-2987
- [9] Thendral T, Vijaya MS, Karpagavalli S, "Discovering Tamil Writer Identity Using Global and Local Features of Offline Handwritten Text", International Review on Computers and Software, Vol.8, N.9, ISSN 1828-6003, September 2013.
- [10] Dunn D, Higgins W.E., "Optimal Gabor filters for texture segmentation", IEEE Transactions on Image Processing, July 1995, ISSN:1057-7149, Page: 947-964.
- [11] F. Shahabi M. Rahmati, "Comparison of Gabor-Based Features for Writer Identification of Farsi/Arabic Handwriting".
- [12] Revathi S V and Vijaya M S, "Predicting the Identity of a Person using Aggregated Features of Handwriting", Advances in Image and Video Processing, Volume 2 No 6, Dec (2014); pp: 23-33
- [13] Zhenyu Heb, Xinge Youb, Yuan Yan Tanga, "Writer identification using global wavelet-based features", Elsevier publication, Neurocomputing (2008), page: 1832-1841.
- [14] Zhenyu He, Yuan Yan Tang, Xinge You, "A Contourlet -based method for writer identification", IEEE International Conference on Systems, Man and Cybernetics, Vol.1, 10-12 Oct 2005, Page: 364-368.
- [15] Marius Bulacu, Lambert Schomaker, Louis Vuurpijl. (2003), "Writer Identification Using Edge-Based Directional Features", Proceedings of the Seventh International Conference on Document Analysis and Recognition.
- [16] Gadkari, Dhanashree, "Image Quality Analysis Using GLCM", UCF Electronic Theses and Dissertations, 2004-12-01.
- [17] Saranya K, Vijaya M.S., "An interactive Tool for Writer Identification based on Offline Text Dependent Approach", International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No. 1, 2013.

- [18] CRAMMER Koby, Yoram Singer, "On the Algorithmic Implementation of Multi-class Kernel-based Vector Machines", Journal of Machine Learning Research, MIT Press, Cambridge, Vol.2 Page 265-292.