A Hybrid Linear Kernel with PCA in SVM Prediction Model of Tamil Writing Pattern

Thendral Tharmalingam *, Vijaya Vijayakumar

Department of Computer Science
PSGR Krishnammal College for Women, Coimbatore, India
*Corresponding author: thendral@psgrkc.ac.in

***Abstract* – Principal Component Regression (PCR) is a regression analysis technique based on Principal Component Analysis (PCA) which enables the identification of the principal components that can be used in a linear kernel and Support Vector Machine (SVM) as a classifier. In PCR, instead of regressing the dependent variable on the explanatory variables directly, the principal components of the explanatory variables are used as regresseors. Only a subset of all the principal components is made use of for regression, thus making PCR a kind of regularized procedure. The principal components with higher variances are selected as regressors and used in SVM linear kernel to estimate the coefficients of the kernel and the linear kernel is specified as Principal Component Kernel–Support Vector Machine (PCK-SVM). Writer Identification in Tamil handwriting is implemented by employing PCK-SVM and the results of the PCK-SVM are compared with our Weighted Least Square regression Kernel based Support Vector Machine (WLK-SVM) and Bayesian linear regression Kernel based Support Vector Machine (BLK-SVM)models. These methods are evaluated on several text images of handwriting at character, word and paragraph levels. The results show that modified linear kernel performs very well with minimum time taken to classify the writer. Performance comparison results of three kernels achieved highest performance of 94.9% accuracy in PCK-SVM than in WLK of 90.8% and BLK of 92.3% accuracy.**

Keywords - Principal Component Analysis, parameter estimation technique, Principal Component Regression, PCK –SVM, Linear kernel, coefficient.

I. INTRODUCTION AND BACKGROUND

Principal Component Analysis (PCA) is a classical statistical method [1-2] for transforming attributes of a dataset into a new set of uncorrelated attributes called Principal Components (PCs). PCA can be used to reduce the dimensionality of a dataset, while still retaining as much of the variability of the dataset as possible. The strength of PCA for data analysis comes from its efficient computational mechanism, the fact that it is well understood, and from its general applicability. PCA is an unsupervised method, which makes no use of information embodied within the class variable.

In most applications, PCA consists of studying p variables measured on n individuals. When n and p are large, the aim is to synthesize the huge quantity of information into an easy and understandable form. Principal component analysis is a normal statistical procedure which has been used to reduce the dimensionality of a dataset [3-4]. It is also known as Karhunen-Loevetrans forms [5]. Some of the applications are data compression, image processing, face recognition, visualization, exploratory data analysis, pattern recognition and time series prediction.

In all of these applications analyzing of the obtained writer identification data becomes complex and challenging task. Since the Tamil handwriting data are usually characterized by different writing styles with much fewer observations, resulting in a high degree of multi-collinearity. To tackle this kind of collinearity problems, latent variable methods, such as Principal Component analysis (PCA) is focused in this work. PCA attempts to find a set of orthogonal principal components (linear combinations of original independent variables) to account for the maximum variations in independent variables.

Principal Component Regression (PCR) is a linear regression model that uses Principal Component Analysis (PCA). PCA is a statistical technique that linearly transforms a data matrix with possibly correlated features into an orthogonal data matrix of uncorrelated features called the principal components [6]. PCR is an ordinary least squares method however, instead of regressing directly on the feature matrix, PCR regresses on the principal components of the feature matrix which is major advantages of PCR is better than the other methods. PCR has a single hyper-parameter, which is the number of components to include in the model.

The proposed modified linear kernel by PCR is introduced for writer identification to identify the writers based on their Tamil handwriting. Parameter estimation method like Weighted Least Square (WLS) [7-8], Bayesian Linear Regression (BLR) [9-10] and Principal Component Regression (PCR) are used to estimate the co-efficient of the linear kernel. These new form of linear kernels with distinctive properties will allow Support Vector Machine (SVM) algorithm[11] to find better optimal hyperplane that discriminates writers in the

feature space. The work was carried out for three types of text like character text, word text and paragraph text by developing three independent datasets.

II. PROPOSED LINEAR KERNEL WITH PCR

The L2 norm SVM formulation is given by,

$$\min_{w,\gamma,\xi} \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2 \quad (1)$$

Subject to

$$d_i (w^T x_i - \gamma) + \xi_i - 1 \geq 0,\ 1 \leq i \leq m,\ \xi_i \geq 0,\ 1 \leq i \leq m \quad (2)$$

The Lagrangian of the objective function [12] is,

$$L(w, \gamma, \xi, u) = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2 - \sum_{i=1}^{m} u_i [d_i (w^T x_i - \gamma) + \xi_i - 1] \quad (3)$$

$$= \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^{m} \xi_i^2 - \left(\sum_{i=1}^{m} u_i d_i x_i^T \right) w - \left(\sum_{i=1}^{m} u_i d_i \right) \gamma - \sum_{i=1}^{m} u_i \xi_i + \sum_{i} u_i \quad (4)$$

Where u are the Lagrangian multipliers. Solving this with Lagrangian duality based on parameters w, γ and ξ the dual problem is obtained as,

$$\max_u L(u) = \sum_{i=1}^{m} u_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} d_i d_j u_i u_j \left(x_i^T x_j + \frac{\delta_{ij}}{C} \right) \quad (5)$$

Subject to

$$\sum_{i=1}^{m} d_i u_i = 0 \quad (6)$$

$$u_i \geq 0, 1 \leq i \leq m \quad (7)$$

The standard form in matrix format is,

$$\min_u L(u) = \frac{1}{2} u^T D (AA^T + \frac{1}{C}) Du - e^T u \quad (8)$$

Subject to

$$d^T u = 0,\ u \geq 0 \quad (9)$$

(Or)

$$\min_u L(u) = \frac{1}{2} u^T Qu - e^T u \text{ Subject to } d^T u = 0,\ u \geq 0 \quad (10)$$

Q can be computed as

$$Q = (A * A^T + I/C) .* (d * d^T) \quad (11)$$

It is noted that the algorithm SVM finally requires three pieces of data Q, d and C where C is the regularization parameter, d is the diagonal matrix of class labels. Q is the obtained from A * A^T and d * d^T[12].

$$AA^T = \begin{pmatrix} x_1^T x_1 & \cdots & x_1^T x_j \\ \vdots & \cdots & \vdots \\ x_i^T x_1 & \cdots & x_i^T x_m \end{pmatrix} = K \quad (12)$$

The i , j^{th} element of AA^T is $x_i^T x_j$ i.e. a dot product of two feature vectors x_i and x_j. The matrix K is called the linear kernel matrix which implies that all information needed for training is captured in the form of dot products of the training vectors. K is positive definite matrix and the set of kernels satisfy closure property. Complex kernels can be defined using simple one and employed in SVM for better learning. Some of the forms of linear kernel K are stated below:

1. $K_1 = A$
2. $K_2 = a.A$
3. $K_3 = AA^T + 1$
4. $K_4 = a.AA^T + 1$

In this work the linear kernel K_2 is used and the parameter 'a' is obtained using parameter estimation method that is Principal Component Regression (PCR). The constant vector 'a' is of dimension equal to number of samples in the training dataset and each element is a parameter added to the sum of the squares of the features.

A. Principal Component Regression

Principal Component Regression (PCR) is a method for evaluating multiple regression data that suffer from multicollinearity. It is a type of disturbance in the data and if present in the data the statistical inferences made about the data may not be reliable. When multicollinearity occurs, least squares estimates are impartial, but their variances are large so they may be far from the true value. It causes multiple errors in the estimation of parameter. Principal components regression reduces the standard errors by adding a degree of bias to the regression estimates.

The PCR method consists of the following three major steps:

1. Perform PCA on the observed data matrix for the explanatory variables to obtain the principal components, and usually select a subset, based on some appropriate criteria, of the principal components so obtained for further use.

2. Now regress the observed vector of outcomes on the selected principal components as covariates, using ordinary least squares regression to get a vector of estimated regression coefficients.

3. Now transform this vector back to the scale of the actual covariates, using the selected PCA loadings the eigenvectors corresponding to the selected principal components to get the final PCR estimator with dimension equal to the total number of covariates for estimating the regression coefficients characterizing the original model.

The classical PCR method as described above is based on classical PCA and considers a linear regression model for predicting the outcome based on the covariates. However, it can be easily generalized to a kernel machine setting whereby the regression function need not necessarily be linear in the covariates, but instead it can belong to the Reproducing Kernel Hilbert Space associated with any arbitrary, symmetric positive-definite kernel.

B. Principal Component Kernel (PCK)

Principal Component Kernel (PCK) is a linear kernel defined using the coefficients derived from PCR, thus facilitating linear kernel based SVM model. PCK is useful when the variance of the feature matrix cannot be well explained with a linear hyperplane. Instead of directly calculating nonlinear principal components, the feature matrix is implicitly mapped into a higher dimensional kernel space where a higher dimensional hyperplane can better fit the direction of highest variance. Therefore, given a p-dimensional random vector $x = (x_1, .. x_p)^T$ with covariance matrix $\sum$ and assume that $\sum$ is positive definite. Let $V = (v_1, \dots v_p)$ be a $(p \times p)$ matrix with orthogonal column vectors that is $v_i^T v_i = 1$ where $i = 1, \dots p$ and $V^T = V^{-1}$. The linear transformation:

$$z = V^T x \tag{13}$$

$$z_i = v_i^T x \tag{14}$$

The variance of the random variable z_i is

$$Var(Z_i) = E[v_i^T x x^T v_i] = v_i^T \sum v_i \tag{15}$$

Maximizing the variance Var (Z_i) under the conditions $v_i^T v_i = 1$ with Lagrange gives

$$\phi_i = v_i^T \sum v_i - a_i (v_i^T v_i - 1) \tag{16}$$

Setting the partial derivation to zero, get

$$\frac{\partial \phi_i}{\partial v_i} = 2 \sum v_i - 2 a_i v_i = 0 \tag{17}$$

Which is

$$\left(\sum - a_i I \right) v_i = 0 \tag{18}$$

In matrix form

$$\left(\sum V \right) = VA \tag{19}$$

of

$$\left(\sum V \right) = VAV^T \tag{20}$$

where $A = diag(a_1, a_2, \cdots, a_p)$. The principal components are orthogonal to all the other principal components since A is a diagonal matrix. Performance of PCK will help to achieve more reliable estimates.

The overall working principle of the proposed system consists of four major steps: The first step is to perform Principal Components Analysis (PCA) on X, using the PCA function, and retaining two principal components. PCR is then just a linear regression of the response variable on those two components. It often makes sense to normalize each variable first by its standard deviation when the variables have very different amounts of variability. Next, fit a PCR model with two principal components. To make the PCR results easier to interpret in terms of the original spectral data, transform to regression coefficients for the original, uncentered variables at step 3. To get the final PCK estimator perform Ordinary Least Squares Regression for estimating the regression coefficients at step 4.

```
Step 1: PCA function
[PCALoadings, PCAScores, PCAVar] = pca
(X,'Economy', false);
Step 2: Perform PCR function
betaPCR = regress(y-mean(y), PCAScores(:,1:2));
Step 3: Perform PCK kernel function
betaPCR = PCALoadings(:,1:2)*betaPCR;
betaPCR = [mean(b1) - mean(K)*betaPCR; betaPCR];
yfitPCR = [ones(length(b1),1) K]*betaPCR;
Step 4: Final OLS regression
b = mvregress(K,b1)
disp('b size');size(b)
For i = 1:size(K,2)
K1(:,i) = (b.*K(:,i));
End
disp ('Kval'); size(K1)
End Process
```

III. EXPERIMENTATION AND RESULTS

In our previous work two other forms of linear kernels were proposed with different parameter estimation techniques. Weighted Least Square (WLS) parameter estimation method was used to estimate the weights for the dot products of the linear kernel WLK. Bayesian linear regression is a parameter estimation method for linear regression in which the statistical analysis is initiated within the outline of Bayesian inference that was used to estimate the coefficients of the linear kernel BLK. This WLK and BLK achieved with distinctive properties allow SVM algorithm to find better optimal hyperplane that discriminates writers in the feature space. SVM with WLK kernel and BLK Kernel has been implemented for three datasets by tuning C- regularization parameter and the predictive accuracies of classifiers are shown in Table I and Table II.

TABLE I. PERFORMANCE OF THE WLK-SVM

Datasets	Accuracy (%)	Precision	Recall	F-measure
Character	72.3	0.733	0.968	0.834
Word	77.2	0.771	0.889	0.826
Paragraph	90.8	0.915	0.8318	0.871

TABLE II. PERFORMANCE OF THE BLK-SVM

Dataset	Accuracy (%)	Precision	Recall	F-measure
Character	74.6	0.812	0.494	0.614
Word	80.1	0.723	0.542	0.619
Paragraph	92.3	0.708	0.952	0.8122

In this proposed work a new kernel called PCK is introduced to identify the writers based on their Tamil handwriting. This new form of linear kernels are defined with the aim of improving the performance of the SVM classifier by adding co-efficient into the dot products of the linear kernel through PCR. The performances of PCK –SVM are analyzed with different types of datasets character (TWINC), word (TWINW) and paragraph (TWINP) and compared with performance of WLK-SVM and BLK-SVM.

The datasets are partitioned into 80% training text images to model the data and 20% of the testing text images to predict the writer. Each type of dataset contains 30000 images of Tamil handwriting with multiclass labels 1 to 300.The profile of dataset is shown in Table 3.The Predictive Accuracy (PA), precision, recall and F-measure are observed for the trained models. Prediction accuracy is the ratio of number of correctly classified instances and the total number of instances. Precision is the segment of retrieved instances that are relevant, recall is the fraction of relevant instances that are retrieved and F-measure computes the average of the information retrieval in precision and recall.

TABLE III. THE PROFILE OF DATASET

Types of dataset	Character	Word	Paragraph
Number of data	30000	30000	30000
Training data	24000	24000	24000
Testing data	6000	6000	6000
Number of features	26	422	422
Number of Class labels	1-300	1-300	1-300

The C-regularization parameters are tuned by using three different values C=1, C=5 and C=10. For example in character type dataset the proposed PCK-SVM achieves 86.6%, 84.2% and 86.6% with C-regularization parameter. SVM with PCK kernel has been implemented for three types of datasets by tuning C- regularization parameter and the predictive accuracies of classifiers are shown in Table IV.

The proposed system achieves 84.1% prediction accuracy in 300 text images of character type dataset. Similarly it produces higher accuracy results for both paragraph and word type dataset. Comparative performances of the three derived kernels are shown in Table 5. The performance comparison of three datasets with training and testing stage results are shown in Table VI.

It is concluded that the proposed PCK-SVM system performs better for all the type of datasets. The overall performance comparison results of three modified linear kernels and linear (lin) kernel are shown in Table VII. However for character type dataset, it produced 70.6%, 72.3%, 74.6% and 86.6% results for linear kernel, WLK-SVM, BLK-SVM and PCK –SVM methods respectively. Table 8 shows the overall performance analysis of proposed PCK-SVM with various types of features.

TABLE IV. SVM WITH PCK KERNEL BY TUNING C- REGULARIZATION PARAMETER

Types of WI	Character			Word			Paragraph		
Parameters	C=1	C=5	C=10	C=1	C=5	C=10	C=1	C=5	C=10
PCK-SVM	86.6	84.2	86.6	85.5	90.4	89.3	91.6	94.9	93.4

TABLE V. COMPARATIVE PERFORMANCES OF THREE DERIVED KERNELS

Parameters	Accuracy(%)		
Types of Writer Identification	WLK-SVM	BLK-SVM	PCK-SVM
Character	72.3	74.6	86.6
Word	77.2	80.1	90.4
Paragraph	90.8	92.3	94.9

TABLE VI. PERFORMANCE ANALYSIS OF PROPOSED PCK WITH VARIOUS TYPES OF FEATURES

Parameters	Class	Training Images	Testing Images	Accuracy (%)	Precision	Recall	F-measure
Character type features	1	24000	6000	86.2	0.80	0.93	0.86
	2	24000	6000	87.8	0.91	0.88	0.90
	3	24000	6000	85.1	0.68	0.78	0.73
	4	24000	6000	89.1	0.73	0.95	0.83
	5	24000	6000	88.6	0.83	0.92	0.87
	. .	. .	. .	. .	. .	. .	. .
	296	24000	6000	84.6	0.75	0.94	0.83
	297	24000	6000	87.1	0.85	0.92	0.88
	298	24000	6000	86.4	0.78	0.96	0.86
	299	24000	6000	87.4	0.80	0.89	0.84
	300	24000	6000	84.1	0.85	0.81	0.83
Word type features	1	24000	6000	87.9	0.85	0.88	0.86
	2	24000	6000	91.8	0.86	0.94	0.90
	3	24000	6000	89.2	0.83	0.92	0.88
	4	24000	6000	91.2	0.85	0.94	0.90
	5	24000	6000	92.6	0.92	0.90	0.91
	. . .	. . .	. . .	. . .	. . .	. . .	. . .
	296	24000	6000	88.9	0.90	0.85	0.87
	297	24000	6000	89.8	0.91	0.86	0.88
	298	24000	6000	89.5	0.90	0.86	0.88
	299	24000	6000	90.9	0.92	0.87	0.89
	300	24000	6000	92.5	0.92	0.90	0.91
Paragraph type features	1	24000	6000	93.9	0.92	0.93	0.92
	2	24000	6000	96.8	0.95	0.97	0.96
	3	24000	6000	92.9	0.89	0.94	0.91
	4	24000	6000	94.5	0.93	0.93	0.93
	5	24000	6000	93.3	0.93	0.91	0.92
	. .	. .	. .	. .	. .	.	. .
	296	24000	6000	95.1	0.92	0.96	0.94
	297	24000	6000	96.3	0.95	0.95	0.95
	298	24000	6000	95.4	0.92	0.96	0.94
	299	24000	6000	96.2	0.93	0.97	0.95
	300	24000	6000	94.8	0.93	0.93	0.93

TABLE VII. OVERALL PERFORMANCE OF THE PCK-SVM KERNEL

Parameters	Accuracy (%)	Precision	Recall	F-measure
Character	86.6	0.872	0.963	0.915
Word	90.4	0.914	0.958	0.935
Paragraph	94.9	0.942	0.976	0.959

TABLE VIII. RELATIVE MEASUREMENTS OF LINEAR AND BLK-SVM KERNEL

Parameters	Accuracy (%)				Precision				Recall				F-measure			
Types of Writer Identification	Lin	WLK-SVM	BLK-SVM	PCK-SVM	Lin	WLK-SVM	BLK-SVM	PCK-SVM	Lin	WLK-SVM	BLK-SVM	PCK-SVM	Lin	WLK-SVM	BLK-SVM	PCK-SVM
Character	70.6	72.3	74.6	86.6	0.689	0.733	0.812	0.872	0.827	0.968	0.494	0.963	0.751	0.834	0.614	0.915
Word	75	77.2	80.1	90.4	0.706	0.771	0.723	0.914	0.748	0.889	0.542	0.958	0.726	0.826	0.619	0.935
Paragraph	88	90.8	92.3	94.9	0.942	0.915	0.708	0.942	0.989	0.831	0.952	0.976	0.964	0.871	0.812	0.959

Receiver Operating Characteristic (ROC) curves for ten output classes are plotted. The more each curve squeezes the left and top edges of the plot, the better the classification. ROC based on precision and recall in character type, word type and paragraph type are depicted in Fig. 1. to Fig. 3.

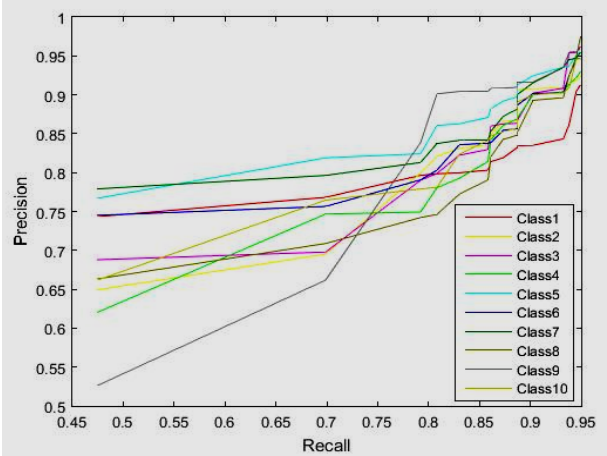

Fig. 1. ROC for character type dataset.

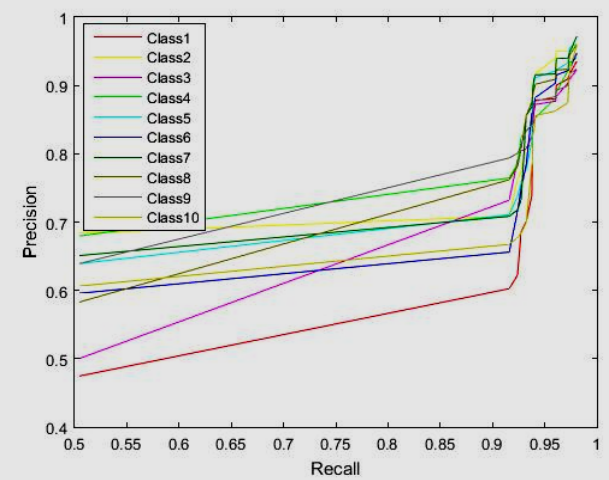

Fig. 2. ROC for word type dataset.

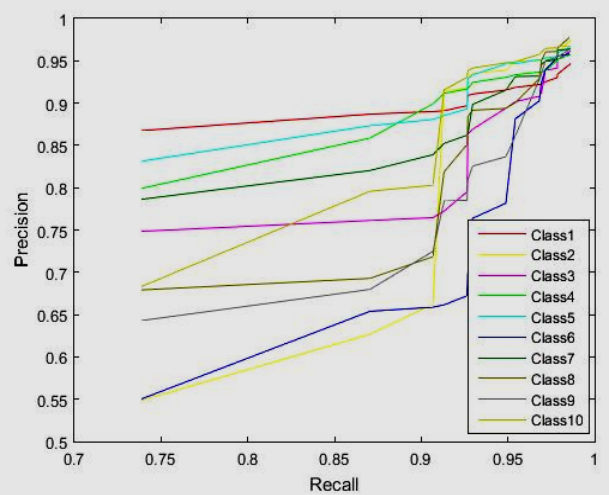

Fig. 3. ROC for paragraph type dataset

It is observed that in the ROC curve for character type, class 8 has high precision of 0.98 and class 9 is curved low at 0.52.In word type, class 7 has high precision of 0.98 and class 1 is curved low at 0.48. In paragraph type, class 8 has high precision of 0.99 and class 2 and class 6 are curved low at 0.52. The performance of principal component regression based SVM prediction models is observed in terms of accuracy for all three datasets and is illustrated in Fig. 4. The comparative performance of WLK-SVM, BLK-SVM and PCK –SVM over linear kernel for character type, word type and paragraph type of writer identification is illustrated in Fig. 5, Fig. 6 and Fig. 7.

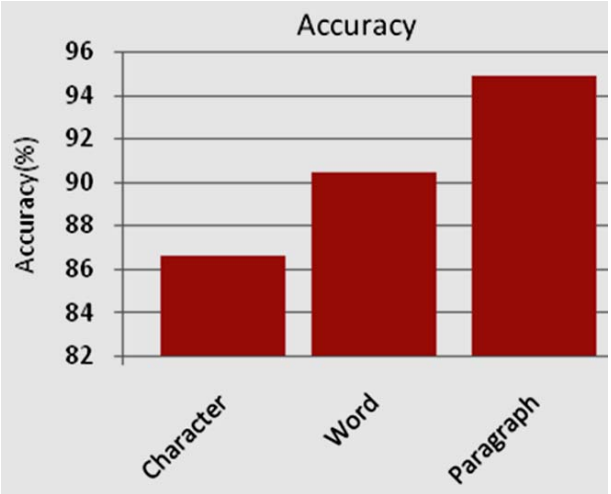

Fig. 4. Prediction Accuracy of PCK.

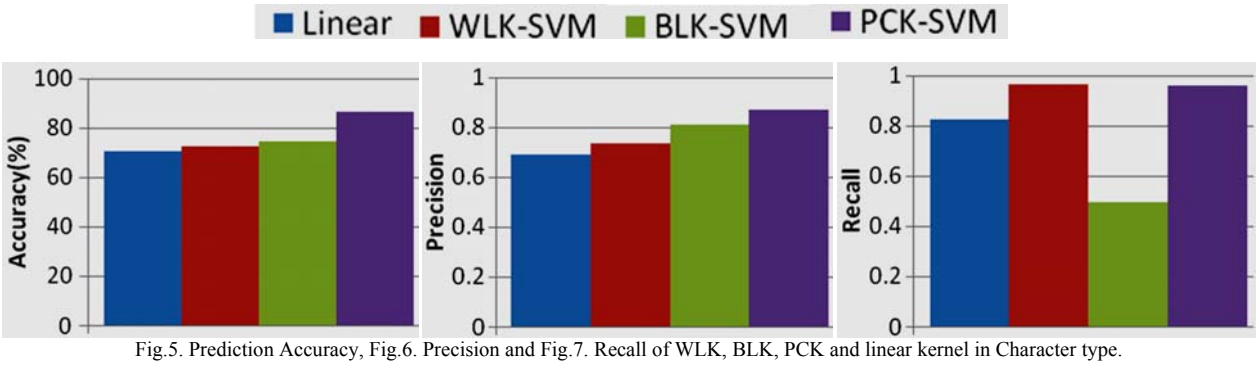

Fig.5. Prediction Accuracy, Fig.6. Precision and Fig.7. Recall of WLK, BLK, PCK and linear kernel in Character type.

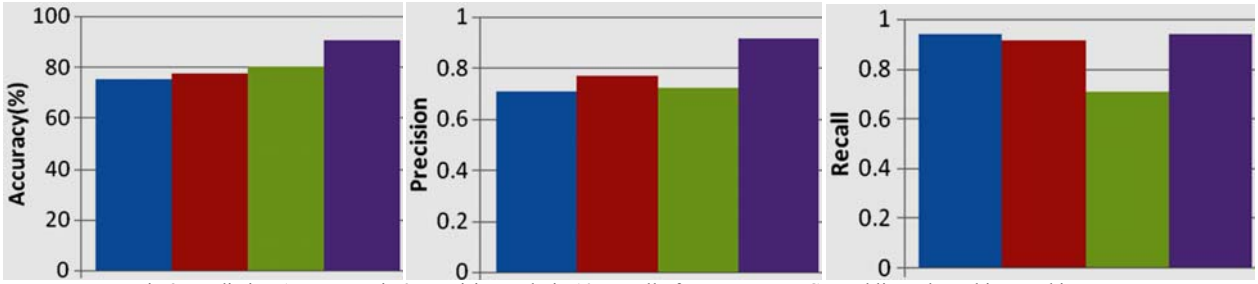

Fig.8. Prediction Accuracy, Fig.9. Precision and Fig.10. Recall of WLK, BLK, PCK and linear kernel in Word images.

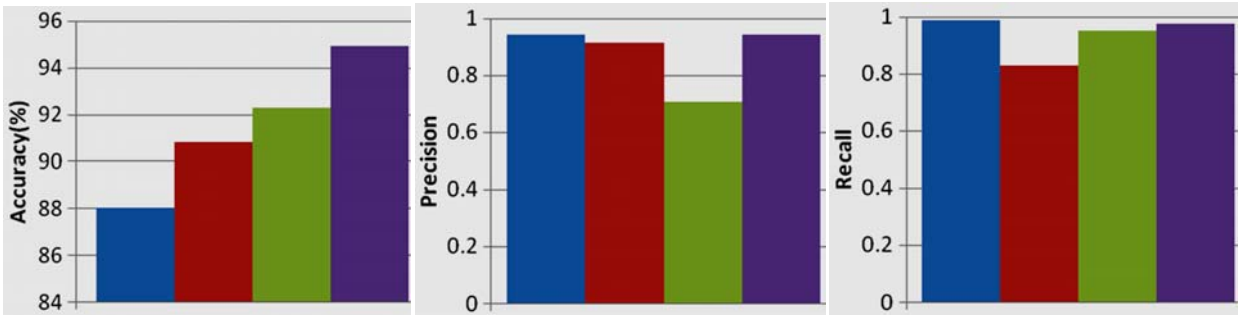

Fig.11. Prediction Accuracy, Fig.12. Precision and Fig.13. Recall of WLK, BLK, PCK and linear kernel in Paragraph images.

A. Findings

Normally Linear kernel computes dot product of features with less computational complexity. The system optimizes only C regularization parameter in linear kernel which makes it faster than other kernels. Linear kernel is found to be best suitable for writer identification if there is huge volume of data with more number of features extracted in it. Hence the new form of linear kernels determines the optimum hyperplane with less computational complexity and achieved better performance. Due to less computational complexity, the time taken to train the model is very less which will help to overcome the situation even if the data is unstructured. Comparative analysis of three newly derived kernels with linear kernel based prediction model shows (94.9%) in paragraph type, 90.4% in word level and 86.6% in character type which confirms high accuracy of PCK-SVM kernel than in linear kernel which confirms (88%) in paragraph type, 75% in word type and 70.6% in character type. The prominent accuracy of 94.9% is achieved using the modified principal component kernel in paragraph images.

IV. CONCLUSION

This paper proposes a Principal Component analysis (PCA) based linear kernel for Support Vector Machine (SVM) based classification of Tamil writing patterns. Principal Component Kernel (PCK) is a variation of Principle Component Regression (PCR) that uses linear kernel. In PCK, new form of linear kernels are defined with the aim of improving the performance of the SVM classifier by adding co-efficient into the dot products of the linear kernel by PCR. Features of Linear kernels are used to enhance its performance by using parameter estimation techniques and the models are built. In PCK, principal component of the feature matrix is a major

advantage which makes it better than the other methods. The proposed new PCK-SVM shows comparatively higher performance when compared to WLK-SVM, BLK-SVM kernels. From the observation it is stated that PCK-SVM achieved better performance with minimum time taken and less computational complexity.

REFERENCES

[1] Holland, S.M., 2008. Principal components analysis (PCA). Department of Geology, University of Georgia, Athens, GA, pp.30602-2501.

[2] Shlens, J., 2014. A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.

[3] Duda (R.), Hart (P.), Stork (D.), Pattern Classification. Second Edition, John Wiley & Sons, Inc., 2001

[4] Agarwal, M., Agrawal, H., Jain, N. and Kumar, M., 2010, February. Face recognition using principle component analysis, eigenface and neural network. International Conference on Signal Acquisition and Processing, 2010. (ICSAP'10), pp. 310-314.

[5] Abdi, H. and Williams, L.J., 2010. Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), pp.433-459.

[6] I.T. Jolliffe, Principal Component Analysis, 2nd edition Springer, 2002.

[7] Shakeeb Khan, Arthur Lewbel, Weighted and Two Stage Least Squares Estimation of Semiparametric Truncated Regression Models, Econometric Theory, 23, 2007, 309–347, May 2003.

[8] Sulaimon Mutiu O., Application of Weighted Least Squares Regression in Forecasting, International Journal of Recent Research in Interdisciplinary Sciences (IJRRIS), Vol. 2, Issue 3, pp: (45-54), July 2015 - September 2015.

[9] Walter, G. and Augustin, T., 2009. Bayesian linear regression.

[10] Chen, T. and Martin, E., 2009. Bayesian linear regression and variable selection for spectroscopic calibration. Analyticachimicaacta, 631(1), pp.13-21.

[11] Chang, C.C. and Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3), 27.1-27.27.

[12] K.P. Soman, R. Loganathan, V. Ajay, Machine Learning with SVM and Other Kernel Methods, PHI Learning Pvt. Ltd., 02-Feb-2009.

[13] B. Scholkopf, A. Smola, K.-R. Müller, Kernel principal component analysis, Advances in Kernel Methods — Support Vector Learning, MIT Press, 1999, pp. 327–352.