CHAPTER-5

PCA BASED GENDER CLASSIFICATION AND WRITER IDENTIFICATION SYSTEM

The proposed work in this chapter is concerned with the development of gender classification and writer identification system based on a very appealing and promising approach i.e., principal component analysis (PCA) which is a competent, proficient, and challenging approach for dimensionality reduction. This chapter describes the introduction, need, mathematical working, merits, and fundamentals of PCA for the proposed experiment in section 5.2. Section 5.3 comprises the PCA-based development of the gender classification system and section 5.4 presents the experimental evaluation of PCA-based development of the writer identification system. Critical and syntactic analysis of the results achieved and the effect of PCA on accuracy and CPU elapsed time has been elaborated in section 5.4.

5.1 INTRODUCTION

Gender classification and writer identification system in Gurumukhi script are the innovative and demanding applications of machine learning that reveal a deep insight into handwriting-based research. After having deep insight into the state-of-the-art work, it has been observed that the development of the gender classification system is a novel attempt with the Gurumukhi script. The promising results revealed always encourage the handwriting-based research communities for the development of many other exigent applications from the handwriting trait, such as age, left or right-handedness, state, nationality, etc.

This chapter covers the PCA based dimensionality reduction method which is an exploratory method to recognize the correlation among the set of variables to achieve the accuracy rate for both gender classification and writer identification by reducing computational time, cost and memory requirements based on locating the principal components.

5.2 PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is a well-known unsupervised dimensionality reduction method which is used to reduce the dimensionality of the large dataset by changing the large set of feature values into smaller ones i.e., eliminating extraneous variables with no loss of information as given by Suri *et al.* (2011), Jolliffe and Cadima (2016) and Das *et al.* (2018). Without losing variability and statistical information, PCA is also known for the orthogonal transformation of data and transform a series of correlated features into a series of no linear relationship.

The first component, i.e., the principal component represents the most variance in the data. These principal components are orthogonal to each other and hence these are statistically independent as shown in Figure 5.1. It is widely used in various applications such as facial recognition, document examination, computer vision, and image compression.

It is quite obvious that with high dimensionality, It is difficult to work because of computational complexity, high memory requirements and challenges to retrieve high accuracy. For dimensionality reduction, we have linear and non-linear methods, as presented by Salem and Hussein (2019). Linear dimensionality reduction methods include PCA, independent component analysis, and orthogonal component analysis. Nonlinear methods for dimensionality reduction are Kernel PCA, local linear embedding (LLE), maximum variance unfolding (Mekala *et al.*, 2019). Khandelwal *et al.* (2016) presented a review on PCA based multimodal biometric systems.

5.2.1 Need of Principal Component Analysis (PCA)

PCA is a widely used technique for dimensionality reduction, developing predictive models, and exploratory data analysis. The most important utilization of PCA is representing multivariable data in a small number of variables by considering trends, jumps, outliers, and clusters. The first principal component has been selected that will maximize the variance of the projected data, Smirg *et al.* (2011). Making the data less interpretable and more independent, PCA has always been a good choice. Another important perspective of PCA is to transform the data in different directions and keeping all the variables in a model for achieving a higher accuracy rate, Kim *et al.*, 2020. PCA is highly recommended with the following paradigms:

- PCA has been used for the best representation of data.
- PCA is a distinguished unsupervised dimensionality reduction method.
- The PCA should always be preferred for the data with high correlation.

- PCA technique is widely used for multi collinear relations that exist between the features and variables.
- PCA worked well with large dimensions of the input features.
- PCA has been a better choice for denoising and data compression.
- For the selection of orthogonal principal components, PCA has been the best choice.

5.2.2 Strengths and weakness with PCA

PCA has numerous benefits when using with the high dimensionality of data. It provides the lower dimension space, thus facilitate the execution of the system with the following benefits.

- Dimensionality Reduction
- Removal of extraneous variables
- Correlated data to un-correlated
- Summarize the information in large data tables by using summary indices.
- Improves performance by reducing over-fitting
- Improves the visualization
- Mathematics of PCA includes findings of the covariance matrix, evaluation of eigenvector
- PC will define more data than others when conversion to orthogonal principal components
- Enhancement in clustering
- Improve text mining
- Basis of multivariate data analysis
- Useful for datasets with imprecise, categorize, missing and multi-colinear data.

Limitations with PCA are:

- New reduced PCA space maximizes the variance of original data
- Work when the relationship between variables is linear
- Interpretation is sensible only when all variables are scaled at the same numerical value
- Lack of probability model
- Independent variables will become less interpretable
- Data standardization is mandatory

- Chances of loss of information
- PCA is a scaled variant
- Hard to work with outliers and missing data

5.2.3 Working of PCA

PCA works on the basis of following step by step execution based on the eigenvalues and the covariance matrix as shown in Figure 5.1.

Step 1: It deals with standardizing the raw data with zero mean and having unit variance.

$$x_j^{(i)} = \frac{x_j^{(i)} - \bar{x}_j}{\sigma_j} \forall_j$$

Step 2: Next step is to generate the covariance matrix of the raw data using.

$$\sum = \frac{1}{m}\sum_{i}^{m}(x_i)\,(x_i)^T, \quad \sum \in R^{n\times n}$$

Step 3: Then compute the eigenvalues and eigenvector of the covariance matrix. Eigenvector has been generated according to eigenvalues in descending order.

$$u^T \sum = \lambda_u$$

$$\boldsymbol{U} = \begin{Bmatrix} | & | & & | \\ \boldsymbol{u_1} & \boldsymbol{u_2} & \dots & \boldsymbol{u_n} \\ | & | & & | \end{Bmatrix}, \boldsymbol{u_i} \in \boldsymbol{R^n}$$

Step 4: Next, select the top k eigenvector of the covariance matrix and that form a novel orthogonal basis for the data.

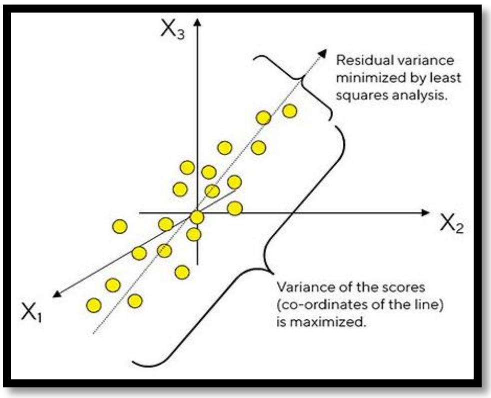

Figure 5.1. PCA based representation of multidimensional variables

5.2.4 Impact of PCA on Feature Vector

As shown in Table 5.1, number of features extracted with Zoning, Diagonal, Transition, and Peak Extent features are 85, 85, 85, and 170 respectively as explained by hierarchical pattern. Also, after hybridization of feature extraction techniques, it has been realized that F1+F2 reveals 85+85=170 values, in the same way, F2+F3 and so on as presented in Table 5.1. Therefore, the goal of the principal component analysis is to reduce the features, i.e., the aim is to select the best principal components. PCA retrieves the least number of features, eliminates insignificant features, and retains generalization as given in Table 5.1. Now the number of features for the Zoning method is 66 which was 85 earlier, in the same way, maximum feature values retrieved by PCA are 278 for F1+F2+F3+F4 which was 425 without implementing PCA.

Table 5.1. PCA based reduction in feature values

Feature Extraction Techniques	Number of features before PCA	PCA based Features
Zoning (F1)	85	66
Diagonal (F2)	85	65
Transition (F3)	85	69
Peak Extents (F4)	170	124
F1+ F2	85+85=170	137
F1+ F3	85+85=170	134
F1+ F4	85+170=255	182
F2+ F3	85+85=170	128
F2+ F4	85+170=255	184
F3+ F4	85+170=255	178
F1+ F2+ F3	85+85+85=255	182
F1+ F2+ F4	85+85+170=340	208
F1+ F3+ F4	85+85+170=340	212
F2+ F3+ F4	85+85+170=340	224
F1+ F2+ F3+ F4	85+85+85+170=425	278

5.3 PRINCIPAL COMPONENT ANALYSIS BASED GENDER CLASSIFICATION SYSTEM

The gender Classification system is a challenging, automated, and efficacious system because of the equivalent visualization of males and females in handwriting. Gender carries rich and distinguished information of an individual characteristic. The present section aims to develop a gender classification system based on PCA i.e., principal component analysis for dimensionality reduction. There are many extraneous variables that can affect the accuracy rate and do not provide any useful information; therefore, dimensionality reduction techniques have been applied for extracting the principal components and improving the accuracy rate thereby lowering the dimensions.

5.3.1 Dataset and Pre-processing

For evaluation of the experiment, the dataset has been prepared from offline handwriting Gurumukhi samples of 200 writers, including 100 male and 100 female writers. This corpus contains in total 70,000 Gurumukhi characters as described in Table 5.2.

Table 5.2. Dataset description for PCA based Gender Classification System

Nature of System	Number of Writers	Number of Classes	Number of Specimens	Total number of Gurumukhi characters
Gender Classification System	100 Females	2	10	35000 female characters
	100 Males	2	10	35000 male characters

For Pre-processing of characters, phases like binarization, slicing, normalization, and thinning were implemented as described in detail in section 3.2.2 The goal of Pre-processing on the data samples is to remove the noise, skeletonization, skewness, and also to eradicate corrupt samples from the dataset. The exclusion of poor-quality data from the datasets helps in improving the gender classification accuracy rate.

5.3.2 Feature Extraction and Classification Techniques

In this sub-section various feature extraction, dimensionality reduction and classification techniques have been executed on the preprocessed characters to attain the results. Table 5.3 shows the summarized view of the approaches used in the proposed experiment on the dataset of 200 writers with 100 male writers and 100 female writers.

Table 5.3. Summarized View of Techniques for Gender Classification System

Feature Extraction Method	Dimensionality Reduction	Classifiers
• Zoning • Transition • Diagonal • Peak Extent Based • Hybridization of feature extraction techniques	• Principal Component Analysis	• Random Forest • Decision Tree

For the proposed experiment, Zoning, Diagonal, Transition, Peak Extent-based feature extraction methods have been implemented, followed by hybridization of feature extraction techniques to generate huge feature values. These values, then become the input to the classification techniques like random forest and decision tree in this experiment.

5.3.3 Experimental Results and Performance Metrics

In this section, experiment results have been discussed in both cases i.e., with and without hybridizing feature extraction techniques and using decision tree and random forest classification techniques. Experiment results like gender classification accuracy along with true positive rate and false positive rate are shown in Table 5.4. The experiment revealed 89.85% gender classification accuracy with F1+F2+F3+F4 and random forest classification technique. In Table 5.5, the result proved satisfactorily that gender classification accuracy has been enhanced and the PCA-based dimensionality reduction technique lowers the dimensions of feature values. After implementing PCA, the maximum gender classification accuracy retrieved is 90.86%, on the process of implementing PCA, CPU elapsed time has been evaluated for gender classification and shown in Table 5.6 and Table 5.7.

Table 5.4. Experimental results for Gender Classification System without PCA

Feature Extraction Techniques	Number of features Without PCA	Decision Tree			Random Forest		
		Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)
Zoning (F1)	85	85.76	81.18	0.41	88.57	87.56	0.41
Diagonal (F2)	85	85.72	81.14	0.31	88.49	88.40	0.30
Transition (F3)	85	84.81	80.63	0.63	88.57	87.50	0.59
Peak Extents (F4)	170	81.76	77.35	0.31	85.52	84.74	0.30
F1+ F2	85+85=170	85.86	81.19	0.52	88.78	87.70	0.50
F1+ F3	85+85=170	84.81	81.45	0.41	89.41	88.34	0.41
F1+ F4	85+170=255	85.12	80.04	0.52	88.37	87.30	0.50
F2+ F3	85+85=170	86.09	81.45	0.41	89.37	88.30	0.41
F2+ F4	85+170=255	84.39	79.97	0.74	88.16	87.10	0.71
F3+ F4	85+170=255	85.29	80.74	0.31	89.06	88.00	0.30
F1+ F2+ F3	85+85+85=255	85.95	81.35	0.21	89.51	88.66	0.19
F1+ F2+ F4	85+85+170=340	85.62	80.66	0.21	88.91	87.72	0.19
F1+ F3+ F4	85+85+170=340	85.78	81.36	0.31	89.55	88.48	0.30
F2+ F3+ F4	85+85+170=340	86.09	81.48	0.21	89.71	88.63	0.19
F1+ F2+ F3+ F4	85+85+85+170=425	85.85	81.25	0.41	**89.85**	**88.75**	**0.41**

Table 5.5. Experimental results for Gender Classification System with PCA

Feature Extraction Techniques	Number of features with PCA	Decision Tree			Random Forest		
		Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)
Zoning (F1)	66	87.83	85.76	0.42	89.56	88.51	0.42
Diagonal (F2)	65	87.79	85.72	0.32	89.48	89.36	0.38
Transition (F3)	69	87.24	84.81	0.64	89.56	88.45	0.58
Peak Extents (F4)	124	83.69	81.76	0.32	86.48	85.66	0.38
F1+ F2	137	87.84	85.86	0.52	89.77	88.66	0.49
F1+ F3	134	88.13	84.81	0.42	90.41	89.3	0.40
F1+ F4	182	86.61	85.12	0.52	89.36	88.25	0.49
F2+ F3	128	88.13	86.09	0.42	90.37	89.26	0.42
F2+ F4	184	86.52	84.39	0.75	89.15	88.05	0.68
F3+ F4	178	87.36	85.29	0.32	90.06	88.96	0.36
F1+ F2+ F3	182	88.02	85.95	0.22	90.51	89.63	0.24
F1+ F2+ F4	208	87.28	85.62	0.22	89.91	88.68	0.24
F1+ F3+ F4	212	88.04	85.78	0.32	90.55	89.44	0.32
F2+ F3+ F4	224	88.17	86.09	0.22	90.71	89.60	0.29
F1+ F2+ F3+ F4	278	87.92	85.85	0.42	**90.86**	**89.72**	**0.42**

Table 5.6. CPU Elapsed Time and Accuracy Rate for Gender Classification without PCA

Feature Extraction Techniques	Number of Features	Random Forest		Decision Trees	
		Accuracy (%)	CPU Elapsed time (ms)	Accuracy (%)	CPU Elapsed time (ms)
Zoning (F1)	85	88.57	9.53	85.76	6.48
Diagonal (F2)	85	88.49	10.5	85.72	7.14
Transition (F3)	85	88.57	10.36	84.81	7.05
Peak Extents (F4)	170	85.52	11.95	81.76	8.13
F1+ F2	85+85=170	88.78	11.82	85.86	8.04
F1+ F3	85+85=170	89.41	11.79	84.81	8.02
F1+ F4	85+170=255	88.37	12.01	85.12	8.17
F2+ F3	85+85=170	89.37	12.36	86.09	8.41
F2+ F4	85+170=255	88.16	14.89	84.39	10.13
F3+ F4	85+170=255	89.06	15.92	85.29	10.83
F1+ F2+ F3	85+85+85=255	89.51	15.01	85.95	10.21
F1+ F2+ F4	85+85+170=340	88.91	17.07	85.62	11.61
F1+ F3+ F4	85+85+170=340	89.55	17.33	85.78	11.79
F2+ F3+ F4	85+85+170=340	89.71	17.58	86.09	11.96
F1+ F2+ F3+ F4	85+85+85+170=425	**89.85**	**19.08**	85.85	12.98

Table 5.7. CPU Elapsed time and Accuracy Rate for Gender Classification with PCA

Feature Extraction Techniques	Number of Features	Random Forest		Decision Tree	
		Accuracy (%)	CPU Elapsed time (ms)	Accuracy (%)	CPU Elapsed time (ms)
Zoning (F1)	66	89.56	8.48	87.83	8.31
Diagonal (F2)	65	89.48	7.14	87.79	7.92
Transition (F3)	69	89.56	7.05	87.24	6.91
Peak Extents (F4)	124	86.48	7.13	83.69	6.99
F1+ F2	137	89.77	7.04	87.84	6.92
F1+ F3	134	90.41	7.02	88.13	6.88
F1+ F4	182	89.36	7.17	86.61	7.03
F2+ F3	128	90.37	8.12	88.13	7.96
F2+ F4	184	89.15	8.14	86.52	7.98
F3+ F4	178	90.06	7.92	87.36	7.76
F1+ F2+ F3	182	90.51	8.16	88.02	8.12
F1+ F2+ F4	208	89.91	7.98	87.28	7.82
F1+ F3+ F4	212	90.55	8.14	88.04	7.98
F2+ F3+ F4	224	90.71	8.27	88.17	7.91
F1+ F2+ F3+ F4	278	**90.86**	**8.95**	87.92	7.79

5.3.4 Syntactic Analysis

To experience the strength of PCA based dimensionality reduction technique for gender classification, we implement feature extraction methods named Zoning, Diagonal, Transition, and Peak Extent and classification has been done by Random Forest and Decision Tree. For the gender classification, with PCA, an accuracy of 90.86% has been attained and without PCA, the accuracy rate was 89.85%. While comparing CPU elapsed time, it was previously 19.08 ms and now the time, it has been reduced to 8.95 ms which is very promising and satisfactory performance as shown in Table 5.8. So, dimensionality reduction for gender classification improves the accuracy from 89.85% to 90.86%, with CPU elapsed time reduced from 19.08 ms to now 8.95 ms.

Table 5.8. Comparative Analysis for Gender Classification

Gender Classification System	Feature extraction and classification techniques used	Accuracy Rate	CPU Elapsed time
Without PCA based implementation	Zoning, Diagonal, Transition, Peak Extent, Random Forest and Decision Tree	89.85%, 88.75% TPR, 0.41% FPR	19.08 ms
With PCA based implementation		**90.86% , 89.72% TPR and 0.42 FPR**	**8.95 ms**

5.4 PRINCIPAL COMPONENT ANALYSIS BASED WRITER IDENTIFICATION SYSTEM

5.4.1 Dataset and Pre-processing

For the evaluation of an experiment, the dataset is prepared from offline handwriting samples of 200 writers in which 100 male and 100 female writers are there. Every writer has written 35 primary characters of the Gurumukhi script. So, the corpus contains 200×35×10=70,000 Gurumukhi characters as depicted in Table 5.9.

Table 5.9. Description of Dataset for Writer Identification

Nature of System	Number of Writers	Number of Classes	Number of Specimen	Total number of Gurumukhi characters
Writer Identification System	200 Writers	200	10	70,000 Gurumukhi characters

5.4.2 Feature Extraction and Classification Techniques

The experimental evaluation for writer identification system consisting of Zoning, diagonal, transition, and Peak Extent based feature extraction methods and for classification, techniques such as decision tree and random forests have been implemented. Hybridization of feature extraction technique has also been implemented. PCA-based dimensionality reduction will reduce the dimensions and compress the data for the reduction in CPU elapsed time too. Table 5.10 shows the feature extraction and classification implemented for the experiment.

Table 5.10. Summarized View of approaches for Writer Identification System

Feature Extraction Method	Dimensionality Reduction	Classifiers
• Zoning • Transition • Diagonal • Peak Extent Based • Hybridization of feature values	• Principal Component Analysis	• Random Forest • Decision Tree

5.4.3 Experimental Results and Performance Metrics

In this section, results for writer identification system based on PCA have been reported. Also, the comparison of identification accuracy, followed by CPU elapsed time is also presented. The accuracy rate for writer identification including, TPR and FPR before and after PCA has been reported in 5.11 and 5.12 and in a similar way, CPU Elapsed time in Table 5.13 and Table 5.14.

Table 5.11. Experimental Results for Writer Identification without PCA

Feature Extraction Techniques	Number of features after PCA	Decision Tree			Random Forest		
		Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)
Zoning (F1)	66	83.19	81.21	0.46	85.91	83.81	0.46
Diagonal (F2)	65	83.15	81.17	0.34	85.84	84.62	0.41
Transition (F3)	69	82.27	80.31	0.69	85.91	83.76	0.62
Peak Extents (F4)	124	79.31	77.42	0.34	82.95	81.11	0.41
F1+ F2	137	83.28	81.30	0.56	86.12	83.95	0.53
F1+ F3	134	82.27	80.31	0.46	86.73	84.56	0.43
F1+ F4	182	82.57	80.60	0.56	85.72	83.56	0.53
F2+ F3	128	83.51	81.52	0.46	86.69	84.52	0.46
F2+ F4	184	81.86	79.91	0.81	85.52	83.38	0.73
F3+ F4	178	82.73	80.76	0.34	86.39	84.24	0.39
F1+ F2+ F3	182	83.37	81.39	0.23	86.82	84.87	0.26
F1+ F2+ F4	208	83.05	81.07	0.23	86.24	83.97	0.26
F1+ F3+ F4	212	83.21	81.23	0.34	86.86	84.7	0.34
F2+ F3+ F4	224	83.51	81.52	0.23	87.02	84.84	0.31
F1+ F2+ F3+ F4	278	83.27	81.29	0.46	87.15	84.96	0.46

Table 5.12. Experimental Results for Writer Identification with PCA

Feature Extraction Techniques	Number of features after PCA	Decision Tree			Random Forest		
		Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)
Zoning (F1)	66	85.20	83.19	0.41	86.87	85.85	0.41
Diagonal (F2)	65	85.16	83.15	0.31	86.8	86.68	0.37
Transition (F3)	69	84.62	82.27	0.62	86.87	85.80	0.56
Peak Extents (F4)	124	81.18	79.31	0.31	83.89	83.09	0.37
F1+ F2	137	85.20	83.28	0.50	87.08	86.00	0.48
F1+ F3	134	85.49	82.27	0.41	87.7	86.62	0.39
F1+ F4	182	84.01	82.57	0.50	86.68	85.60	0.48
F2+ F3	128	85.49	83.51	0.41	87.66	86.58	0.41
F2+ F4	184	83.92	81.86	0.73	86.48	85.41	0.66
F3+ F4	178	84.74	82.73	0.31	87.36	86.29	0.35
F1+ F2+ F3	182	85.38	83.37	0.21	87.79	86.94	0.23
F1+ F2+ F4	208	84.66	83.05	0.21	87.21	86.02	0.23
F1+ F3+ F4	212	85.40	83.21	0.31	87.83	86.76	0.31
F2+ F3+ F4	224	85.52	83.51	0.21	87.99	86.91	0.28
F1+ F2+ F3+ F4	278	85.28	83.27	0.41	**88.13**	**87.03**	**0.41**

Table 5.13. CPU Elapsed time and Accuracy Rate for Writer Identification without PCA

Feature Extraction Techniques	Number of Features	Random Forest		Decision Trees	
		Accuracy (%)	CPU Elapsed time (ms)	Accuracy (%)	CPU Elapsed time (ms)
Zoning (F1)	85	85.91	8.77	83.19	5.96
Diagonal (F2)	85	85.84	9.66	83.15	6.57
Transition (F3)	85	85.91	9.53	82.27	6.49
Peak Extents (F4)	170	82.95	10.99	79.31	7.48
F1+ F2	85+85=170	86.12	10.87	83.28	7.4
F1+ F3	85+85=170	86.73	10.85	82.27	7.38
F1+ F4	85+170=255	85.72	11.05	82.57	7.52
F2+ F3	85+85=170	86.69	11.37	83.51	7.74
F2+ F4	85+170=255	85.52	13.7	81.86	9.32
F3+ F4	85+170=255	86.39	14.65	82.73	9.96
F1+ F2+ F3	85+85+85=255	86.82	13.81	83.37	9.39
F1+ F2+ F4	85+85+170=340	86.24	15.7	83.05	10.68
F1+ F3+ F4	85+85+170=340	86.86	15.94	83.21	10.85
F2+ F3+ F4	85+85+170=340	87.02	16.17	83.51	11
F1+ F2+ F3+ F4	85+85+85+170=425	**87.15**	**17.55**	83.27	11.94

Table 5.14. CPU Elapse time and Accuracy Rate for Writer Identification with PCA

Feature Extraction Techniques	Number of Features	Random Forest		Decision Trees	
		Accuracy (%)	CPU Elapsed time (ms)	Accuracy (%)	CPU Elapsed time (ms)
Zoning (F1)	85	86.87	7.8	85.2	7.65
Diagonal (F2)	85	86.8	6.57	85.16	7.29
Transition (F3)	85	86.87	6.49	84.62	6.36
Peak Extents (F4)	170	83.89	6.56	81.18	6.43
F1+ F2	85+85=170	87.08	6.48	85.2	6.37
F1+ F3	85+85=170	87.7	6.46	85.49	6.33
F1+ F4	85+170=255	86.68	6.6	84.01	6.47
F2+ F3	85+85=170	87.66	7.47	85.49	7.32
F2+ F4	85+170=255	86.48	7.49	83.92	7.34
F3+ F4	85+170=255	87.36	7.29	84.74	7.14
F1+ F2+ F3	85+85+85=255	87.79	7.51	85.38	7.47
F1+ F2+ F4	85+85+170=340	87.21	7.34	84.66	7.19
F1+ F3+ F4	85+85+170=340	87.83	7.49	85.4	7.34
F2+ F3+ F4	85+85+170=340	87.99	7.61	85.52	7.28
F1+ F2+ F3+ F4	85+85+85+170=425	**88.13**	**8.23**	85.28	7.17

5.4.4 Syntactic Analysis

For the experimental evaluations for writer identification system using Zoning, Diagonal, Transition, and Peak Extent based features, it has been revealed that without using PCA based techniques, accuracy and CPU time reported was 87.15% and 17.55 ms. And after implementing the strengths of the PCA-based technique, now the accuracy and time have been improved to 88.13% and 8.23 ms as shown in Table 5.15.

Table 5.15. Syntactic Analysis for Writer Identification Accuracy

Writer Identification System	Techniques Used	Accuracy Rate	CPU Elapsed time
Without PCA	Zoning, Diagonal, Transition, Peak Extent, Random Forest and Decision Tree	87.15%.	17.55 ms
With PCA		**88.13%, 87.03% TPR, 041% FPR**	8.23 ms

5.5 CONCLUSION

In this chapter experimental work has been performed for the development of a gender classification system and writer identification system based on the principal component analysis. Without implementing PCA, the gender classification accuracy of 90.86% has been experienced with CPU elapsed time of 8.95 ms and before PCA, classification accuracy was 89.85% in 19.08 ms. Similarly, without implementing PCA, we experience maximum writer identification accuracy of 87.15% with a CPU elapsed time of 17.55 ms. After reducing the features with PCA, improved accuracy of 88.13% has been revealed with 8.23 ms. Thus, with PCA, the target of achieving high accuracy results has been successfully achieved.