CHAPTER-4

GENDER CLASSIFICATION AND WRITER IDENTIFICATION SYSTEM USING HYBRIDIZATION OF VARIOUS FEATURE EXTRACTION TECHNIQUES

The development of gender classification and writer identification systems based on offline handwritten text in the Gurumukhi script is an incredible and challenging task. This chapter aims to present the experiment evaluation by implementing feature extraction techniques, namely, Zoning, Diagonal, Transition, and Peak Extent and for the classification, three classification techniques, namely, k-NN, Decision Tree, Artificial neural network (ANN), Multi-layer perceptron (MLP), and Random Forest (RF) have been implemented on offline handwritten Gurumukhi samples. To experience a high accuracy rate, we have implemented a novel feature extraction method based on the hybridization of zoning, diagonal, transition and peak extent techniques. This chapter is divided into five main sections. Section 4.1 deals with the introduction, section 4.2 describes the successful implementation of feature extraction techniques followed by section 4.3 that discusses the strength of the hybridizing feature extraction techniques and their novel application to the proposed system. Section 4.4 comprises the experiment results for gender classification with hybridization of feature extraction techniques. Section 4.5 consists of the experiment evaluation for writer identification using hybridization of feature extraction techniques followed by the discussion and conclusion in section 4.6.

4.1 INTRODUCTION

The aspiration of gender classification and writer identification system is to assign the gender and the authorship of the handwritten text to a definite author, out of numerous authors in the stored dataset. Handwriting is an expression of an individual that cannot be imitated and is an efficient attribute in representing thoughts and ideas. On the basis of offline handwritten samples, many innovative applications have been evolving by researchers. Chaudhari and Thakkar (2019) presented a deep survey on handwriting traits and discussed the effects of handwriting traits on the personality and psychological aspects.

The implementation of the proposed experiment consists of two main methods. In the first method, implementation of feature extraction techniques has been implemented successively with the classification techniques and in the second method, an advanced approach i.e., hybridization or fusion of feature extraction methods have been accomplished to achieve successful results followed by the comparison of the results.

4.2 FEATURE EXTRACTION TECHNIQUES

For the experimental evaluation, different feature extraction methods like zoning, diagonal, transition, and peak extent have been implemented. The working scenario of these feature extraction methods has been deeply elucidated in section 3.2.3. And, after the implementation of feature extraction techniques, number of feature values are extracted which are represented in Table 4.1. Here we see that 85 feature values have been be extracted from the zoning method for every character. Similarly, 85 feature values can be extracted from the diagonal method for every character and so on.

Table 4.1. Number of features extracted with feature extraction methods

Feature Extraction Techniques	Number of Features
Zoning (F1)	85
Diagonal (F2)	85
Transition (F3)	85
Peak Extents (F4)	170

4.3 HYBRIDIZATION OF FEATURE EXTRACTION TECHNIQUES

Hybridization of feature extraction techniques aims at combining the strengths of the feature extraction methods for enhancing the accuracy of results. This section is concerned with the implementation of hybridization of feature extraction techniques and achieves a novel contribution that allows feature extraction techniques to be hybridized in all possible different combinations to draw the most discriminative features.

Pradeep *et al.* (2012) presented hybridization of features with salient features of K-NN, backpropagation RNN recurrent, neural network, neural network, and radial basis function to recognize English character. Khanduja *et al.* (2015) developed a hybrid feature and classifier approach for the Devanagari script and achieved 93.4% results. Goel and Vishwakarma (2016) also brought forward the hybridization of

Discrete wavelet transform and discrete cosine transform with SVM to achieve the desired promising results. Katiyar and Mehfuz (2016) proposed identification of handwritten character by hybridization of feature extraction methods on centre of excellence for document analysis and recognition, (CEDAR) data samples. Shaikh *et al.* (2016) presented hybridization of auto-learned features (ALF) and human-engineered features (HEF) for handwriting verification with CNN and Autoencoder (AE) and produces a maximum of 99.7% accuracy. Singh and Singh (2019) brought forward a hybrid method that helps in integrating the complementary strength of feature extraction techniques and concluded that feature weighting wrapper method, the extended adjusted ratio of ratio proves to be the best method.

For instance, in Table 4.2, F1+F2 i.e., hybridizing zoning and diagonal feature extraction techniques generates 170 features that have been acting as input to classifiers, in the same way, F1+F4 generates 85+170 = 255 feature values and F1+F2+F3+F4 produces a maximum of 425 feature values. Table 4.2 shows a maximum of 425 feature values after hybridization of all four feature extraction techniques, F1+F2+F3+F4 whereas the maximum feature value that has been obtained without hybridization is 170, as shown in Table 4.1.

Table 4.2. Number of Features extracted with hybridization of feature extraction techniques

Feature Extraction Techniques	Number of Features
Zoning (F1)	85
Diagonal (F2)	85
Transition (F3)	85
Peak Extents (F4)	170
F1+ F2	85+85=170
F1+ F3	85+85=170
F1+ F4	85+170=255
F2+ F3	85+85=170
F2+ F4	85+170=255
F3+ F4	85+170=255
F1+ F2+ F3	85+85+85=255
F1+ F2+ F4	85+85+170=340
F1+ F3+ F4	85+85+170=340
F2+ F3+ F4	85+85+170=340
F1+ F2+ F3+ F4	85+85+85+170=425

4.4 GENDER CLASSIFICATION SYSTEM BASED ON HYBRIDIZATION OF FEATURES EXTRACTION TECHNIQUES

In this section, the development of a novel feature extraction method i.e., feature extraction method based on the hybridization or fusion of feature extraction techniques is presented to combine the dominant features. The collection of data samples, Pre-processing, feature extraction, hybridization of feature extraction, and classification are presented and discussed. The experimental results retrieved and the comparative analysis are presented in a tabular and diagrammatic representation in the following subsections.

4.4.1 Dataset and Pre-processing

The first and foundational step is the collection of data samples. Here, for the implementation of the desired objective, a dataset is generated consisting of scanned images of offline handwritten Gurumukhi characters collected from 150 writers with 75 female writers and 75 male writers, and each person has generated 10 copies of 35 primary characters of Gurumukhi script. Thus, the total data sample collection is $150 \times 35 \times 10 = 52{,}500$ Gurumukhi characters. Both datasets i.e., the female dataset consists of $75 \times 35 \times 10 = 26{,}250$ Gurumukhi characters and the male dataset contains $75 \times 35 \times 10 = 26{,}250$ Gurumukhi characters. The female and male dataset must be stored separately and hence generating a two-class membership problem.

To start with the implementation process, first, it is mandatory to convert raw data into digitized data using scanning, which is done at 300 dpi (dots per inch), a standard value for scanning. Pre-processing deals with cleaning and maintaining the data i.e., removing noise, slicing, normalizing, and thinning of the characters. The Pre-processing phase consists of the following major sub-phases: -

- Binarization of the data means setting threshold values that are transformed of the grey level image from 0-255 spectrum into 0-1 spectrum.
- Slicing data deals with cutting of characters from the document and cropping whitespaces.

- Normalization will convert all characters into the window of 64×64 using the nearest neighborhood interpolation algorithm (NNI), followed by generating bmp images with setting height, width, color, depth ratio, cropping, and auto-filling.
- Last but not least, implementing parallel thinning algorithm and the generation of the thinned image is an important phase and is the primary requirement of feature extraction methodologies as shown in Figure 3.4.

4.4.2 Feature Extraction and Classification Techniques

Feature extraction methodologies will extract hidden features of writers from their offline handwriting samples of characters. For the proposed experimental task, Zoning, Diagonal, Transition, and Peak Extent based features have been implemented on the datasets. Preprocessed Gurumukhi characters of female and male writers acted as input to feature extraction methods that produced feature values corresponding to these methods followed by implementation of classification techniques k-NN, Decision Tree, Random Forest, and Adaptive Boosting which were discussed in detail in section 3.2.5. Results obtained after individual implementation of feature extraction and classification techniques are shown in Table 4.3.

Table 4.3. Gender Classification accuracy with feature extraction techniques

Feature Extraction Technique	K-NN (C1) (%)	Decision Tree (C2) (%)	Random Forest (C3) (%)	Adaptive Boosting(C4) (%)
Zoning (F1)	57.3	58.9	60.2	69.3
Diagonal (F2)	52.2	51.5	53.6	63.8
Transition (F3)	61.2	62.1	65.2	77.1
Peak Extent (F4)	53.8	57.1	61.9	69.4

Here, the maximum accuracy achieved for gender classification is 77.1% with the features extracted from the Transition method and classification by the Adaptive Boosting method.

Next, we have implemented hybridization of feature extraction techniques i.e., implementing hybridization of zoning, diagonal, transition, and peak extent-based feature extraction methods in all possible ways, and then classification techniques are applied to retrieve the best gender classification result. Finally results obtained are given in Table 4.4 and highlighted results shows the maximum accuracy of 94.6% using hybridization of all feature extraction methods with the adaptive boosting classifier.

Table 4.4. Gender Classification Accuracy with hybridization of feature extraction techniques

Feature Extraction /Classification Techniques	K-NN (C1) (%)	Decision Tree (C2) (%)	Random Forest (C3) (%)	Adaptive Boosting (C4) (%)
Zoning (F1)	57.3	58.9	60.2	69.3
Diagonal (F2)	52.2	51.5	53.6	63.8
Transition (F3)	61.2	62.1	65.2	77.1
Peak Extent (F4)	53.8	57.1	61.9	69.4
F1+F2	61.3	61.8	63.7	74.5
F1+F3	67.5	69.0	71.5	83.4
F1+F4	61.7	64.4	67.8	77.0
F2+F3	64.6	64.8	67.7	80.3
F2+F4	59.4	60.8	64.7	74.6
F3+F4	66.7	69.1	73.7	85.0
F1+F2+F3	71.7	72.5	75.2	88.3
F1+F3+F4	73.5	76.0	79.9	92.1
F1+F2+F4	72.9	74.8	78.5	90.5
F2+F3+F4	74.7	76.2	80.7	93.9
F1+F2+F3+F4	91.7	92.9	92.8	94.6

4.4.3 Experimental Results and Performance Metrics

After implementing the hybridization of feature extraction techniques, we have seen a tremendous improvement from 77.1 to 94.6 % which is a great and remarkable endeavour for the novel application. In this section, performance metrics that have been evaluated to study the strength of the experiment have been discussed and the results of these parameters such as precision, false-positive rate, root mean square error, the area under curve are shown in graphical representations in the next sub-sections. Figure 4.1 shows the gender classification accuracy with respect to four classifiers as K-NN, Decision Tree, Random Forest, and Adaptive Boosting and four feature extraction techniques as presented in Table 4.1.

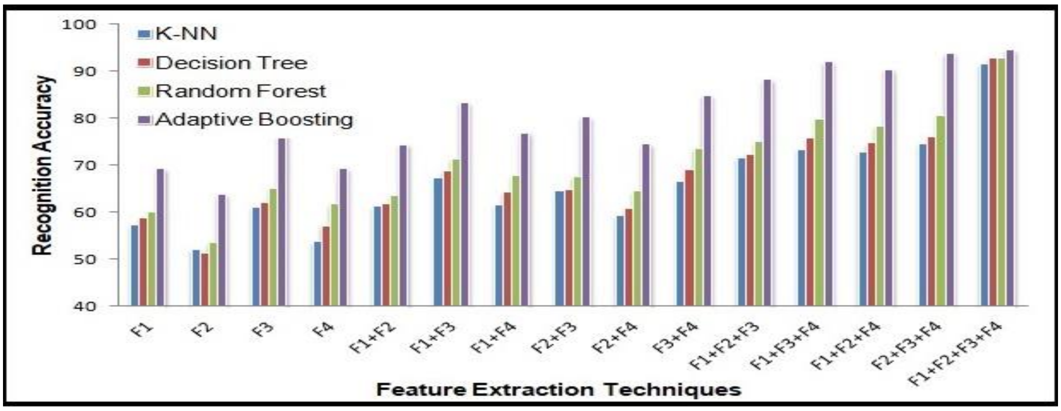

Figure 4.1. Gender Classification Accuracy with hybridization of Feature Extraction Techniques

4.4.3.1 Accuracy Rate

It is defined as the measure of the success of the system. Accuracy is defined as how perfectly results are achieved i.e., percentage of correct predictions of test data. Accuracy can be calculated by dividing the number of correct predictions retrieved divided by the total number of predictions. By determining the accuracy rate, we can predict the best classifier or best combination of feature extraction methods with the classification methods.

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

Where TP and TN stand for true positive and true negative and FP and FN stand for False positive and false negative. Thus, experiments revealed 94.6% as shown in Table 4.4.

4.4.3.2 False Positive Rate

It is defined as the ratio of a number of false positives to the sum of false-positive and true negatives. It is the proportion of all negatives that still yield positive test outcomes and is also defined as the probability of falsely rejecting the null hypothesis. It is an incorrect identification of anomalous data.

$$\text{False Positive Rate} = \frac{FP}{FP + TN}$$

A false-positive rate is also called a false alarm rate and is defined as the rate at which a positive result will be generated when the true value is negative. The current experiment revealed a maximum accuracy rate of 94.6% with FPR 2.0% as shown in Table 4.5.

Table 4.5. False Positive Rate for gender classification

Features	K-NN (C1) (%)	Decision Tree (C2) (%)	Random Forest (C3) (%)	Adaptive Boosting (C4) (%)
Zoning (F1)	5.3	5.2	5.0	4.7
Diagonal (F2)	5.3	5.7	5.2	4.9
Transition (F3)	4.5	4.2	3.8	3.5
Peak Extent (F4)	4.9	4.8	4.4	4.2
F1+F2	5.1	5.2	4.9	4.6
F1+F3	4.5	4.3	4.0	3.8
F1+F4	4.8	4.8	4.5	4.2
F2+F3	4.6	4.7	4.2	3.9
F2+F4	4.8	4.9	4.5	4.3
F3+F4	4.5	4.3	3.9	3.7
F1+F2+F3	4.3	4.3	4.0	3.8
F1+F3+F4	4.3	4.1	3.8	3.6
F1+F2+F4	4.4	4.5	4.2	4.0
F2+F3+F4	4.2	4.2	3.8	3.6
F1+F2+F3+F4	2.2	2.2	2.1	2.0

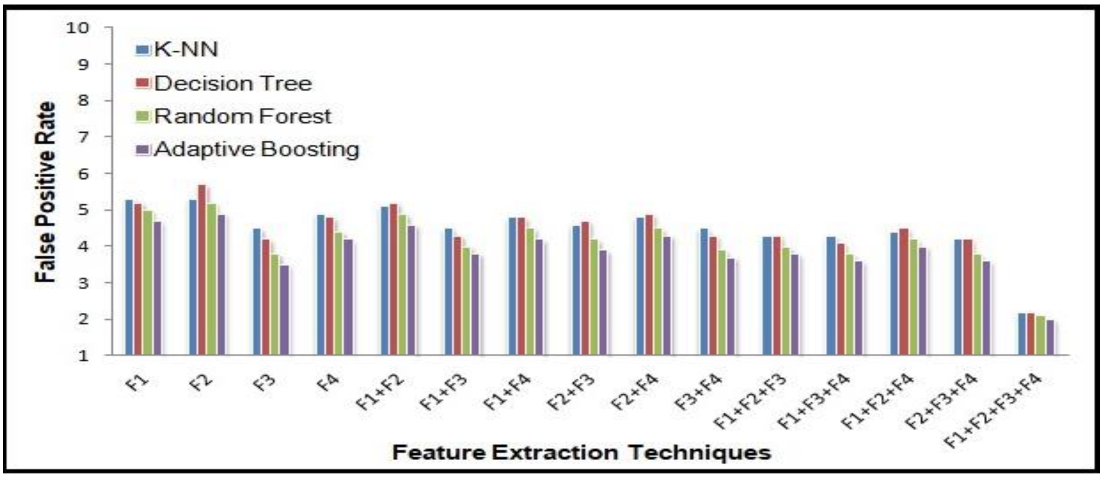

Figure 4.2. False Positive Rate

4.4.3.3 Precision Rate (PR)

Precision proves the number of positive class predictions that will actually belong to the positive class. So, it is the fraction of relevant instances to the total retrieved instances. Table 4.6 depicts the precision rate of 94.4%. These results are graphically shown in Figure 4.3.

Table 4.6. Precision Rate for gender classification

Features	K-NN (C1) (%)	Decision Tree (C2) (%)	Random Forest (C3) (%)	Adaptive Boosting (C4) (%)
Zoning (F1)	54.9	58.2	58.9	67.6
Diagonal (F2)	48.9	49.4	51.3	61.7
Transition (F3)	58.7	62.3	64.9	75.6
Peak Extent (F4)	50.9	56.9	61.9	68.1
F1+F2	58.1	60.3	61.7	72.4
F1+F3	64.8	68.7	70.6	81.6
F1+F4	58.7	63.9	67.0	75.3
F2+F3	61.3	63.7	66.2	78.3
F2+F4	55.9	59.5	63.4	72.7
F3+F4	63.6	69.1	73.5	83.3
F1+F2+F3	68.3	71.4	73.5	86.1
F1+F3+F4	70.2	75.7	79.2	90.2
F1+F2+F4	69.1	73.5	76.9	88.2
F2+F3+F4	70.8	75.3	79.6	91.7
F1+F2+F3+F4	92.0	92.5	92.5	94.4

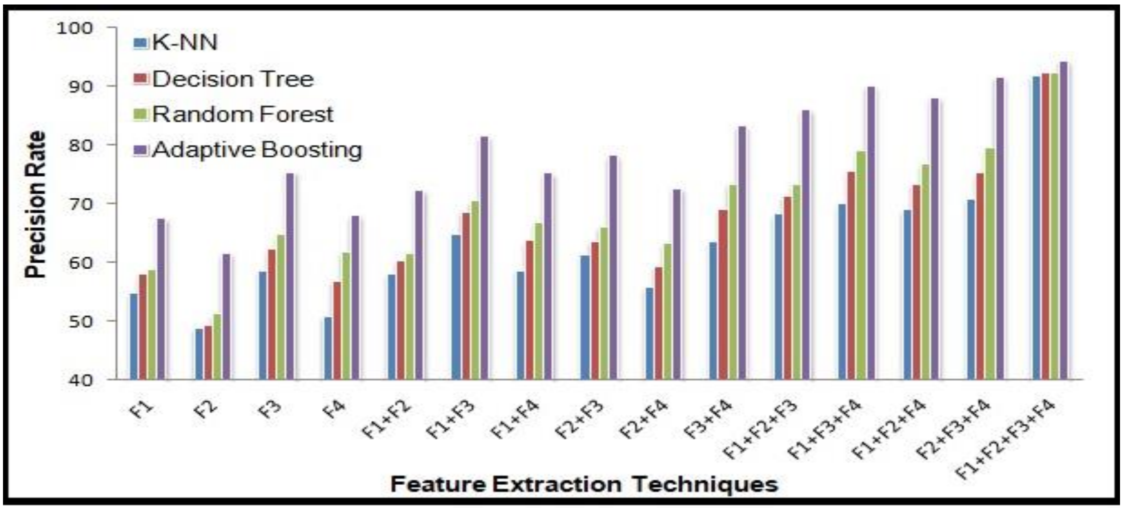

Figure 4.3. Precision Rate

4.4.3.4 Root Mean Square Error (RMSE)

Root mean squared error tells the concentration of the data variables around the line of best fit. It is the standard deviation of the prediction error and is also calculated as the difference between the value obtained and the value observed. Table 4.7, shows the values of RMSE for the proposed gender classification system with different feature values and classifiers. RMSE values for different feature extraction and classifiers for the proposed experiment are graphically depicted in Figure 4.4.

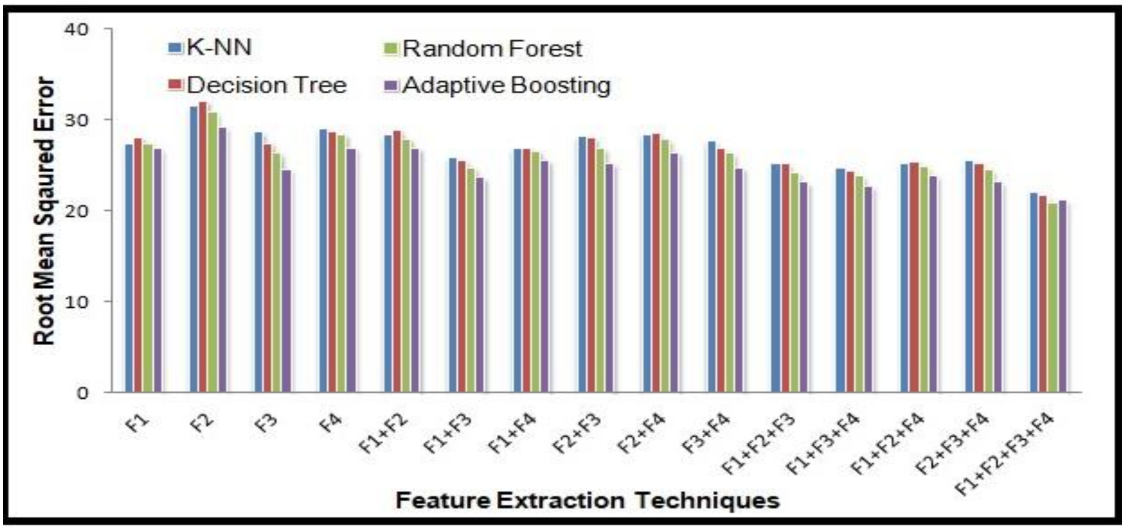

Figure 4.4. Root Mean Squared Error

Table 4.7. Root Mean Square Error for gender classification

Features	K-NN (C1) (%)	Decision Tree (C2) (%)	Random Forest (C3) (%)	Adaptive Boosting (C4) (%)
Zoning (F1)	27.5	28.1	27.5	26.9
Diagonal (F2)	31.6	32.2	30.9	29.3
Transition (F3)	28.7	27.5	26.4	24.6
Peak Extent (F4)	29.1	28.7	28.5	27.0
F1+F2	28.4	28.9	28.0	27.0
F1+F3	25.9	25.6	24.8	23.7
F1+F4	26.9	27.0	26.6	25.6
F2+F3	28.3	28.1	26.9	25.3
F2+F4	28.5	28.6	27.9	26.5
F3+F4	27.7	27.0	26.4	24.8
F1+F2+F3	25.2	25.2	24.3	23.2
F1+F3+F4	24.7	24.4	23.9	22.8
F1+F2+F4	25.3	25.5	24.9	23.9
F2+F3+F4	25.6	25.3	24.6	23.2
F1+F2+F3+F4	22.1	21.7	20.9	21.2

4.4.3.5 Area under Curve (AUC)

It is used to measure the quality of the classification models and is defined as the definite integrals between two points.

$$AUC = \int_a^b f(x)dx$$

AUC of a classifier is defined as the chance that the classifier will position an indiscriminately selected positive example higher than a randomly chosen negative example. Table 4.8 shows the values of the strength of the curve *i.e.*, area under curve with the graphical view in Figure 4.5.

4.4.4 Syntactic Analysis

After realizing the experimental results, it has been analysed that the effect of hybridization of feature extraction techniques really boosts the accuracy rate in the novel experiment, from 77.1% to 94.6%, which is really a satisfactory and remarkable achievement. This means that experiment with hybridization of feature extraction has efficiently produced high accuracy in comparison to the successful implementation of feature extraction techniques. It is concluded that for the proposed experiment, hybridization has facilitated with the better results in this experiment. The results achieved in both the cases, i.e., without and with hybridization of feature extraction techniques are shown in Table 4.9.

Table 4.8. Area Under Curve for gender classification

Features	K-NN (C1)	Decision Tree (C2)	Random Forest (C3)	Adaptive Boosting (C4)
Zoning (F1	52.5	52.6	52.8	82.9
Diagonal(F2)	50.6	50.7	50.8	80.3
Transition(F3)	54.1	54.1	54.2	86.7
Peak Extent(F4)	53.3	53.2	53.2	83.2
F1+F2	57.7	57.8	58.0	88.7
F1+F3	60.8	60.8	61.0	89.9
F1+F4	58.7	58.7	58.8	90.0
F2+F3	59.7	59.7	59.9	89.8
F2+F4	58.2	58.2	58.2	91.1
F3+F4	62.3	62.2	62.3	91.2
F1+F2+F3	66.0	66.1	66.3	90.4
F1+F3+F4	68.2	68.2	68.4	91.8
F1+F2+F4	69.9	69.9	70.0	91.2
F2+F3+F4	70.6	70.6	70.7	92.4
F1+F2+F3+F4	93.7	94.3	94.2	**95.1**

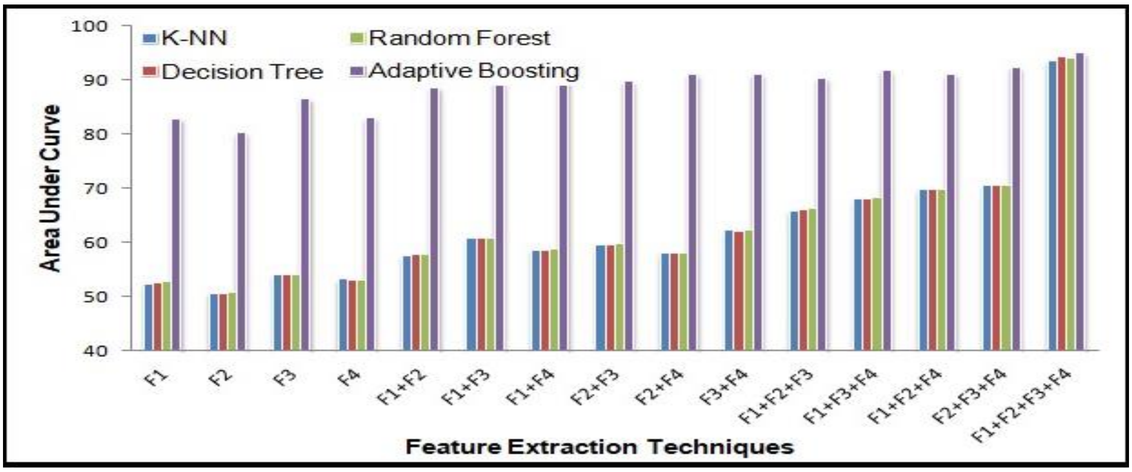

Figure 4.5. Area Under Curve

Table 4.9. Comparison of Gender Classification Accuracy

Gender Classification System	Dataset	Techniques	Results
Before implementing hybridization of feature extraction technique	150 writers with 75 female writers and 75 male writers	Implementing Transition Feature Extraction method and Adaptive Boosting classifier	77.1%
After implementing hybridization of feature extraction technique F1+F2+F3+F4	150 writers with 75 female writers and 75 male writers	Implementing Zoning, Diagonal, Transition and Peak Extent F1+F2+F3+F4 and Adaptive Boosting classifier	**94.6% Accuracy, 94.4% Precision and 2.0% FPR**

4.5 WRITER IDENTIFICATION SYSTEM BASED ON HYBRIDIZATION OF FEATURE EXTRACTION TECHNIQUES

4.5.1 Dataset and Pre-processing

To develop a writer identification system, based on offline handwriting samples, a dataset comprising offline handwritten samples from 150 writers has been generated, so the dataset for the proposed experiment consisting of a total of 150×35×10 equals 52,500 Gurumukhi characters.

Then after collecting data samples from the writers, scanning was done at 300 dpi (dots per inch). Next, is the Pre-processing phase, which includes binarization, then setting of threshold values between black and white pixels to [0,1] followed by slicing of characters and then normalization and generation of bitmap images. Normalization helps in providing uniformity and confined it to a specific window of size 64×64 followed by thinning which means reducing the width of the character from several pixels to a single pixel. Thinned images are the images on which the feature extraction methods will be executed.

4.5.2 Feature Extraction and Classification Techniques

To achieve the successful results for writer identification, we implemented Zoning, Diagonal, Transition, and Peak Extent Based feature extraction techniques and for classification, Random Forest, Artificial neural network, and multi-layer perceptron techniques were executed. So let us discuss the strength of these classification methods.

Artificial Neural Network (ANN) (C1) classifier is a computing system that is inspired by biological neurons. It is used for many application areas because of its self-learning capability, less complexity, the requirement of fewer parameters, backpropagation, learning, and reprogramming. As the problem of misspecifications is less in ANN so, it is considered as a suitable tool best for handwriting-based researches and is also called a universal approximator. Secondly, Multi-Layer Perceptron Model (MLP) (C2) has been exploited for its full connectedness property. It has many layers which are best suited for the non-linearly separable data and are sensitive with hyperparameter and with feature scaling, thus sometimes poses inconveniency. It is a supplement of the feed-forward neural network, consisting of the input layer, hidden layers, and output layer, and is trained using a backpropagation algorithm.

Random Forest (RF) (C3) classifier is an ensemble learning algorithm that works on decision trees and used for classification and regression methods. It uses feature randomness and bagging and is also named as meta estimator that will average the performance of decision trees to improve the accuracy and also to control overfitting. It takes less training time and generates high accuracy even when some data is missing. RF has many applications in banking, marketing, medicine, land usage and less suitable for regression-based applications. Maximum identification accuracy of 92.05% is obtained with TPR 91.97% and FPR 0.39% without implementing hybridization of feature extraction techniques as shown in Table 4.10.

Then again next step is to perform the hybridization of the feature extraction technique, and the results achieved are shown in Table 4.11. Feature extraction techniques are hybridized to produce a feature vector with more features obviously, thus achieving a maximum accuracy rate of 93.36% which is better than the previous results. The goal of hybridization of feature extraction techniques is to explore maximum features and to form a large-size feature vector. Fusion of feature extraction techniques will reveal all possible features associated with the character.

4.5.3 Experimental Results and Performance Measures

For the evaluation of the algorithm, three performance parameters such as accuracy, true positive rate, and false-positive rate have been evaluated. As shown in Table 4.10, maximum writer identification accuracy of 92.05% has been realized with the Zoning and Random Forest classification technique.

Table 4.10. Writer Identification Accuracy Rate without hybridization of feature Extraction techniques

Feature Extraction Technique	ANN Classifier			MLP Classifier			Random Forest Classifier		
	Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)
Zoning (F1)	86.57	85.44	0.39	87.05	85.91	0.39	92.05	91.97	0.39
Diagonal (F2)	86.53	85.40	0.39	87.01	85.87	0.30	91.95	92.85	0.29
Transition (F3)	85.99	84.49	0.49	86.47	84.96	0.59	92.03	91.90	0.58
Peak Extent (F4)	82.49	81.44	0.20	82.95	81.90	0.30	88.87	89.00	0.29

After getting the results in Table 4.10, we experience the accuracy rates with hybridizing approach and we observe that after implementing the new technique, the accuracy rate has been improved from 92.05% to 93.36%.

4.5.3.1 Writer Identification Accuracy

The graphical view of writer identification accuracy is shown in Figure 4.6. The reported result outperforms in comparison to the literature survey as discussed in Sections 2.2 and 2.3 and also poses futuristic and upcoming directions to the researchers such as age, gender, handedness, physiological autopsy, personality, stress, and even nationality identification based on handwriting. Maximum identification accuracy of 93.36% when using F1+F2+F3+F4 with Random Forest classifier (C3) is reported as presented in Table 4.11. Results obtained for each classifier are:

- ✓ 86.90% with F2+F3+F4 and ANN Classifier
- ✓ 87.38% with F2+F3+F4 and MLP Classifier
- ✓ 93.36% with F1+F2+F3+F4 and Random Forest Classifier

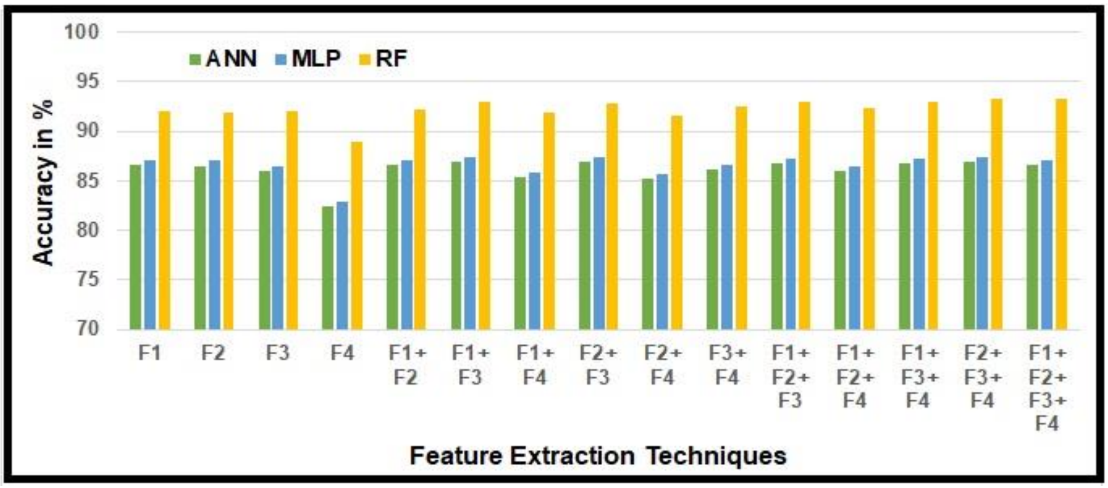

Figure 4.6. Writer Identification Accuracy with hybridization of feature Extraction techniques

Table 4.11. Writer Identification Accuracy Rate with hybridization of feature extraction techniques

Feature Extraction Techniques	ANN Classifier			MLP Classifier			Random Forest Classifier		
	Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)	Accuracy (%)	TPR (%)	FPR (%)
Zoning (F1)	86.57	85.44	0.39	87.05	85.91	0.39	92.05	91.97	0.39
Diagonal (F2)	86.53	85.40	0.39	87.01	85.87	0.30	91.95	92.85	0.29
Transition (F3)	85.99	84.49	0.49	86.47	84.96	0.59	92.03	91.90	0.58
Peak Extent (F4)	82.49	81.44	0.20	82.95	81.90	0.30	88.87	89.00	0.29
F1+ F2	86.58	85.54	0.59	87.06	86.01	0.49	92.25	92.12	0.49
F1+ F3	86.87	84.49	0.39	87.35	84.96	0.39	92.91	92.78	0.39
F1+ F4	85.36	84.81	0.39	85.84	85.28	0.49	91.83	91.69	0.49
F2+ F3	86.87	85.76	0.69	87.35	86.24	0.39	92.87	92.74	0.39
F2+ F4	85.27	84.07	0.30	85.75	84.54	0.69	91.61	91.48	0.68
F3+ F4	86.09	84.97	0.39	86.58	85.44	0.30	92.55	92.43	0.29
F1+ F2+ F3	86.75	85.62	0.79	87.24	86.10	0.20	93.01	93.13	0.19
F1+ F2+ F4	86.02	85.29	0.49	86.50	85.77	0.20	92.39	92.14	0.19
F1+ F3+ F4	86.77	85.47	0.59	87.26	85.94	0.30	93.05	92.92	0.29
F2+ F3+ F4	86.90	85.76	0.39	87.38	86.24	0.20	93.22	93.09	0.19
F1+ F2+ F3+ F4	86.65	85.53	0.69	87.14	86.00	0.39	**93.36**	**93.23**	**0.39**

4.5.3.2 TRUE POSITIVE RATE

It is also called as sensitivity or recall, which means the probability that actual positive samples will test positive i.e., accurately predicted.

$$\text{True Positive Rate} = \frac{TP}{TP + FN}$$

i.e., the ratio of a number of true positives to the sum of a number of true positives and false negatives. For our current experiment, with respect to a maximum accuracy rate of 93.36%, the value of TPR achieved is 93.23%. Performance evaluation metrics named TPR also called true positive rate which means to calculate how many correct positive results are there among all positive samples. The value of TPR is as shown in Table 4.11 and Figure 4.7.

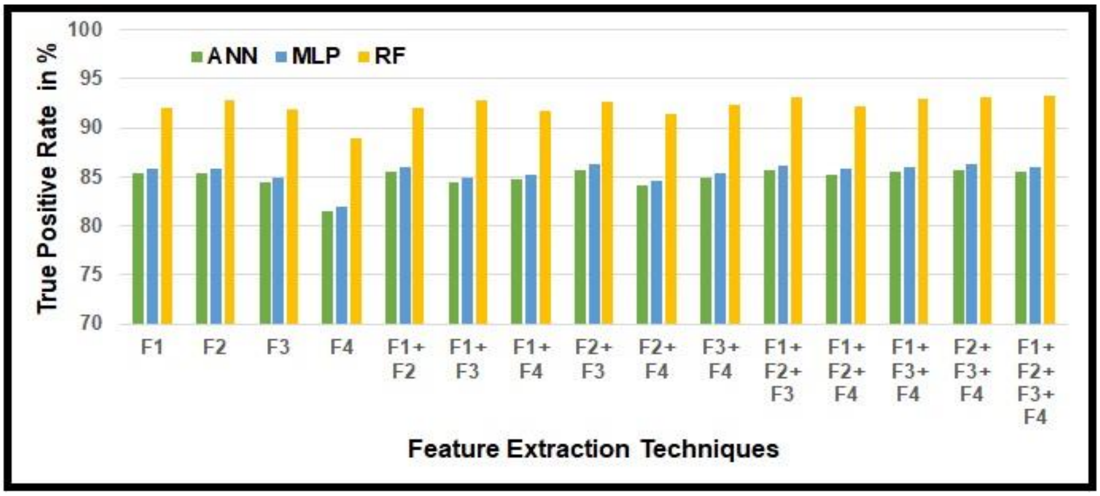

Figure 4.7. True Positive Rate for writer identification

4.5.3.3 FALSE POSITIVE RATE

False-positive rate means how many incorrect positive outcomes occurred during negative samples available during the testing of experimental results. The value of FPR during the evaluation of writer identification accuracy is as shown in Table 4.11 and Figure 4.8 which is 0.39% for the maximum accuracy of 93.36%.

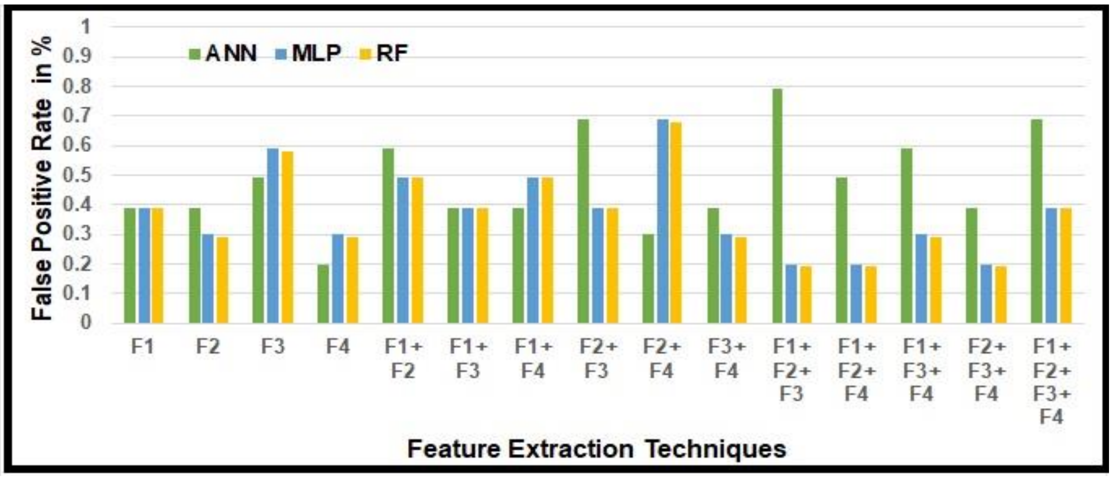

Figure 4.8. False Positive Rate for writer identification

4.5.4 Syntactic Analysis

It has been analyzed from the above experimental findings that the accuracy rate for writer identification without implementation of hybridization and using the hybridizing approach has been improved from 92.05% to 93.36% as shown in Table 4.12.

Table 4.12. Comparison of Writer Identification Accuracy

Writer Identification System	Dataset	Techniques	Results
Before hybridization of feature extraction techniques	150 writers, 52500 samples	Zoning Feature extraction method with random forest classification technique	92.05% Accuracy, 91.07% TPR and 0.39 FPR
After implementing hybridization of feature extraction techniques, i.e., F1+F2+F3+F4	150 writers, 52500 samples	Zoning, Diagonal, Transition and Peak Extent Based with random forest classification technique	**93.36% Accuracy, 93.23% TPR and 0.39% FPR**

4.6 DISCUSSION AND CONCLUSION

The development of gender classification and writer identification system based on offline handwriting in Gurumukhi script with the implementation of feature extraction techniques followed by the development of a novel method based on hybridization of feature extraction techniques has been presented in this chapter. It is a constructive and incredible application for forensic investigations, forgery detection, identifying suspects, questioned documents, etc. Diagonal, Zoning, Transition, Peak Extent-based features have implemented for extracting features, and hybridization of feature extraction methods has also been experienced in this chapter in many different ways for boosting up the accuracy rate. For classification, we used Decision Tree, Random Forest, K-NN, Adaptive Boosting for gender classification, and ANN, MLP, and RF for writer identification. Experimental results revealed the maximum gender classification accuracy based upon hybridization of feature extraction techniques is 94.6% with Zoning, Diagonal, Transition and Peak Extent based Adaptive Boosting classification technique with Precision 94.4%, area under curve, root mean square error, FPR 2.0%. For writer identification, accuracy of 93.36% has been achieved with Diagonal, Zoning, Transition, Peak Extent based features and Random forest classifier along with 93.23% TPR and 0.39% FPR. The result obtained for gender classification and writer identification has remarkably improved with hybridization and hence presented a distinguish direction to the researchers as compared to the state of the art work for boosting the results.