International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

Handwritten Devanagari Text Recognition using Single Classifier Approach with VSPCA Scheme

Vijay More[1], Prashant Yawalkar[2], Madan Kharat[3]
[1]Dept. of Computer Engineering
MET's Institute of Engineering, Bhujbal Knowledge City
Nashik, Affiliated to SPPU, Pune, India
vbmore2005@rediffmail.com
[2]Dept. of Computer Engineering
MET's Institute of Engineering, Bhujbal Knowledge City
Nashik, Affiliated to SPPU, Pune, India
prashant25yawalkar@gmail.com
[3]HOD, Dept. of Computer Engineering
MET's Institute of Engineering, Bhujbal Knowledge City
Nashik, Affiliated to SPPU, Pune, India
mukharat@rediffmail.com

Abstract— In this research paper we used individual classifier approach for Handwritten Devanagari text recognition. We experimented different categorical classifiers namely Random Forest Classifier (RFC), Support Vector Machine (SVM), K Nearest Neighbor Classifier (KNN), Logistic Regression Classifier (LogRegr), Decision Tree Classifier (DTree). Seven different feature sets are used namely Eccentricity, Euler Number, Horizontal Histogram, Vertical Histogram, HOG Features, LBP Features, and Statistical Features. The experimentation is carried out on 9434 different characters whose features are extracted from 220 handwritten image documents from PHDIndic_11 dataset. We deduced and implemented a unique scheme namely VSPCA scheme. VSPCA is Vectorization, Scaling, and Principal Component Analysis carried out on all feature sets before being given for model training. We obtained varied accuracies using all these five classifiers on all these six feature sets in which 99.52% highest accuracy is observed.

Keywords- Classifier, Devanagari, features, handwritten documents, text recognition, training.

I. INTRODUCTION

National language of India is Hindi and in Maharashtra state Marathi is the official language. These languages and several other languages viz. Konkani, Sanskrit, Maithili, Bhojpuri and Nepali are composed using Devanagari script. As a mother language, these are spoken from childhood and taught to write from preliminary schooling. English is third language in many states of India where Hindi is not a key regional language.

Marathi, as Devanagari script, is an Indo-Aryan language which is predominantly spoken by people in the Maharashtra state of India. However, it is co-official language in Goa state and the India's union territory of Daman, Diu & Silvassa (sourced in Wikipedia). Generally Marathi language is also spoken in border regions of states whose border is attached with Maharashtra.

Devanagari script has long historical existence which is used for writing literature, trade, domestic communication, official communication using letters, books, religious Holy books like Geeta, Ramayana and literature from various Saints in Maharashtra. Devanagari languages are also used in politics. All these aspects are motivating factors to work on Devanagari script language documents to identify handwritten text.

Many researchers tried to put their efforts on recognition of unconstrained handwritten text, written in Devanagari script. The objective behind their effort mostly aligned to explore more complexities embedded in Devanagari script. Since, Devanagari script has its own inherent complexities that cover more phonemes of language constructs. Majority of Indian languages having some sort of curvature, upper modifier, lower modifier, middle zone, many conjunctives, composite characters, shirorekha, vertical bar, and many more different components which are not seen in other foreign language scripts like Roman, and this is the main reason it makes very complex to segment and recognize Devanagari text.

Researchers' community attracts such type of challenges to address them enthusiastically and try to overcome any issues while performing operations during experimentation.

This paper is organized in various sections and each one is discussed separately. Section II is dedicated to address literature survey, section III deals with proposed method, experimental setup and dataset, section IV focuses on results

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

and discussion, section V highlights on conclusion and future scope and lastly references mentioned.

II. LITERATURE SURVEY

During last four decades majority of the work is carried out on recognition of printed text in initial three decades. Handwritten text recognition is seen addressed in more sense during last decade only. At best of our knowledge, handwritten Devanagari text recognition is addressed by few researchers only.

In a research [1], the author used abstract mathematical features, structure based features and script dependent features with total 41 features given to series of Multi Layer Perceptron network having 16 hidden layers and 6 output layers. The author obtained 92.8% accuracy after the experimentation and highlighted to use more features to make the system generalize.

Handwritten script identification is carried out using Directional Discrete Cosine Transform D-DCT given to KNN and LDA classifiers in the work [2] carried out by the author and obtained accuracy of 96.95% and 85.77% using KNN and LDA respectively by experimenting on 12 features over different script documents including Devanagari.

The work proposed by author [3] for character recognition using MLP classifier which is two-layer feed forward neural network with back-propagation learning. Number of neurons used by author was 70 and 40 in the first and second hidden layers. Author carried out experimentation on CVPR dataset with extracting different kinds of features and obtained accuracy of 93.4%.

Using fractal features, component based feature and topological features, author [4] used MLP classifier to recognize script from handwritten Devanagari post cards and obtained 89.48% accuracy.

Author [5] extracted 192 co-occurance histogram based texture features using 2D-DWT Haar wavelet and performed classification using KNN classifier. Author used 50-50% samples for training and testing and obtained 97.5% average classification accuracy for single writer document for different scripts.

Handwritten script recognition for different 7 Indic scripts carried out by author [6] using DCT and Wavelet features at block level. Author used 60% blocks for training and 40% blocks for testing. Classification is done on Kannada, English, Hindi, Malayalam, Punjabi, Tamil, Gujarati, and Telagu language scripts using KNN classifier and obtained average accuracy of 96.4%.

In other work [7], author extracted 13 stroke features from 100 Devanagari script documents which are given to KNN classifier and obtained 92% average accuracy for Devanagari script in combination with Roman and Urdu.

Some two-stage work carried out by author in [8] proposing classification of Devanagari words from Indian documents containing English, Devanagari and Bengali scripts. Using SVM classifier, author obtained 98.51% accuracy in their work.

Author in [9] worked for handwritten post cards to recognize Bangla, English and Devanagari scripts. Author extracted Using fractal features, busy-zone features and topological features from words segmented and are given to MLP classifier for recognition. Author claimed 96.79% accuracy in their work.

Using 3 layer MLP classifier, author [10] worked to recognize handwritten Devanagari text. A separate MLP is used for 32 intersection features, 16 shadow feature, 200 chain code histogram features and 48 straight line fitting features. Classification is carried out using voting classifier with MLP for separate feature set and obtained accuracy is 92.80%.

Author in [11] extracted global features like stroke density and local features like aspect ratio, eccentricity, extent which are extracted from word images of IAM database. Author performed classification using KNN classifier on Hindi words in combination with Roman and Kannada and obtained accuracy of 96.05%.

Author in [12] suggested that global approaches were more inclined to generalize for script identification than local approaches. Author extracted energy profile of a script texture using log-Gabor filters in first level in hierarchical model as coarse features, and Ratios of normalized energies and horizontal profile in second level of hierarchy as finer features. These features are tested using KNN and Parzen classifiers and accuracy obtained is 94%.

There was a survey among different features and classifiers carried out by the author of this paper in [13] and subsequent segmentation techniques are discussed in [14] and [15]. PHDIndic_11 dataset [16] is used for experimentation of this research paper.

III. PROPOSED METHOD AND EXPERIMENTAL SETUP

Proposed methodology is given in Fig. 1.

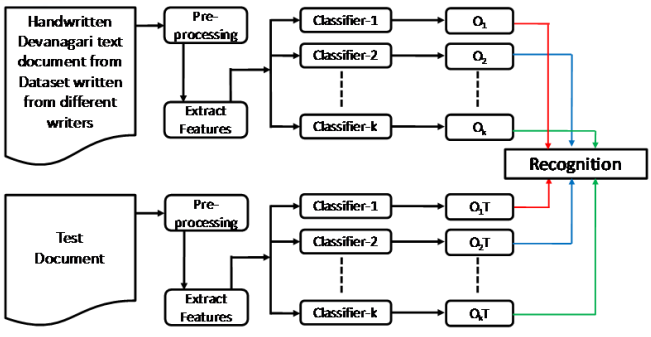

Figure 1. Proposed Methodology

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

Components of the Proposed Methodology

- Pre-processing
- Feature extraction
- Feature selection
- Appropriate Model selection
- Model training, and
- Recognition

A. Pre-Processing

This part deals with the image segmentation using statistics inherent within handwritten characters. Every handwritten character is segmented and stored as separate image in a folder with unique name. All these characters embedded in document are collected and stored in the folder. Following activities in pre-processing are carried out:

- Converting image to gray level using thresholding method
- Image dilation
- Image erosion
- Image Transformations
- Image Filtering
- Segmentation

B. Feature extraction

After segmentation, features are extracted which represent the sense of information. These features play major role in pattern recognition. Following feature groups are extracted:

- Eccentricity
- Euler Number
- Horizontal Histogram
- Vertical Histogram
- Histograms of Oriented Gradients
- LBP
- Statistical Features

Features are the sense of information which was extracted from the input scanned and cleaned image. All this information must be similar for similar images but must not be same for other images. These features play major role in pattern recognition. Thus the selection of feature extraction technique becomes a key aspect in achieving high recognition performance. In other words, this stage is used to remove or reduce redundancies (unwanted information) from data. Difference between feature extraction and feature selection must be clarified before processing. Feature extraction is a technique used to extract unique features from the input image [18]. On the other hand, feature selection is the process of selection of most relevant features from available pool of features which helps to improve the classification accuracy of the extracted features [19]. Extracted features list is depicted in (TABLE I.).

TABLE I. FEATURES EXTRACTED

Feature name	Number of features
Eccentricity	1 feature x 9434 images
Euler Number	2 features x 9434 images
Horizontal Histogram	20 features x 9434 images
Vertical Histogram	20 features x 9434 images
HOG Features	400 features x 9434 images
LBP Features	60 features x 9434 images
Statistical Features	110 features x 9434 images
Total	613 features x 9434 images

C. Feature Selection

Following steps carried out for feature selection:

- Vectorization
- Feature Scaling
- Principal Component Analysis

Vectorization is a process where all the target labels i.e. character ids, are taken into separate column and indicated as "1" in place wherever it is applicable and "0" in place where it is not applicable. The vector for each target label is appended along with existing features. In short, label vectors acting as additional feature which is purely independent and not correlated to other label vectors. Independent features are more important in the process of recognition.

Feature Scaling is a process where magnitudes of all the features are scaled to a single common scale. This is required because if the magnitude of one feature is larger, say 100 to 1000, and magnitude of other feature is varying between 1 to 10, then higher magnitude feature dominates smaller magnitude feature and biased result may be generated that in turn hamper the recognition result. To cope-up with this situation, feature scaling is needed. There are different types of scaling techniques available among which standard scalar is considered in this work and the formula is given in (1).

$$x_{new} = \frac{x_i - \bar{x}}{\sigma} \quad (1)$$

Here x_{new} is scaled observation from original observation x_i for which $\bar{x}$ is mean and σ is standard deviation.

Principal Component Analysis (PCA) is a process where most significant features are gathering at the start of feature vector and least important features are at the end. PCA is performed on existing feature set after converting vectorization and feature scaling. PCA also helps to reduce the dimensionality of feature set and taking initial subset of features as principal components which hold most significant information within them. These most significant features are used for recognition of text in further process.

Steps for PCA Computation:

- *Standardization*: dataset standardized (normalized) to make all the features in same scale which helps

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

classifier not to confuse with different magnitude valued features.

- *Covariance Matrix Computation*: It is used to understand how features are varying from mean with respect to each other to identify any relationship between them.
- *Identify the principal components*: Compute the Eigen vectors and Eigen values of the covariance matrix to identify the principal components. Magnitude of Eigen value gives percentage of information contained in principal components.
- *Feature vector selection*: We keep all Eigen vectors obtained or discard those of lesser significance (low Eigen values), and form with the resultant matrix of vectors that we call Feature vector.
- *Final Dataset Computation*: Align the data from the original axes represented by the principal components using feature vector formed by the Eigen vectors of the covariance matrix. This can be done by multiplying the transpose of the original data set by the transpose of the feature vector.

Vectorization, Scaling, and Principal Component Analysis is VSPCA scheme that we are proposing by which our system performed better than the methodologies suggested earlier.

D. Appropriate model selection

Since, every character represents its own identity, the features extracted belongs to specific character. This means that our dataset is categorical dataset and requires training models which deals with categories. Therefore the categorical models are considered for experimentation on these features. These models are:

- Random Forest Classifier
- Support Vector Machine
- K Nearest Neighbor Classifier
- Logistic Regression Classifier
- Decision Tree Classifier

These models are considered in single classifier approach as well as hybrid approach with voting classifier. In single classifier approach, every model is trained separately with features having target label and tested on unseen data (i.e. by removing target label, the features are given to model) to recognize the label of feature. In hybrid approach, voting classifier is given a set of different classifier as a base classifier and majority voting is carried out. For Eg: 3 classifier identified character k as k and 2 identified as f, then 3 classifiers who identified k as k are considered and set of rules generated by voting classifier is built using set of rules followed by 3 classifiers. This makes voting classifier robust with all the labels, which in turn helps to identify a character more correctly.

E. Model training

Every model considered here is trained with 80% training data (with labels) and 20% testing data (without labels) [17] from existing feature groups. All these feature sets used for model training. During model training classifier specific parameters are tuned for every classifier to obtain good training results so that during testing phase, the model performs well and recognize the character with highest accuracy.

F. Recognition

Recognition of text i.e. identifying label for features, is carried after rigorous training phase. Unseen data is given to the trained model and character label is identified which is nothing but the identification of text from handwritten Devanagari character.

E.g. Images of the characters क, ख, ग, घ must be recognized by the system as क, ख, ग, घ characters respectively. The model should not identify other character in place of intended character. The model which does this recognition task with very less error, yields acceptable accuracy.

G. Experimental Setup

Experimentation is carried out on 64bit Windows running on computer system comprising processor of Intel(R) Core(TM) i3-3217U CPU having 1.80 GHz speed with 4GB RAM. Python 3.9.12 compiler is used for compilation of all the programs.

H. Dataset Information

PHDIndic_11 [16] is a page-level handwritten document image dataset of 11 official Indic scripts for script identification.

These scripts are: Bangla, Devanagari, Roman, Urdu, Oriya, Gurumukhi, Gujarati, Tamil, Telugu, Malayalam and Kannada. Among them Devanagari script of 220 different documents written by different authors are considered in this work. Different categories of Devanagari page level images are observed by author [14].

IV. RESULTS AND DISCUSSION

In case of data pre-processing, all standard practices are experimented. VSPCA scheme is adopted in this work which is less seen much in other research.

VSPCA scheme experimented on LBP features, the performance of different classifiers is seen in (TABLE II.).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

TABLE II. VSPCA PERFORMANCE WITH LBP FEATURES

Technique[a]	Time (sec)	RFC	SVM	KNN	Log Regr	D-Tree
-V,-S,-PCA	1019.14	48.86%	30.76%	35.08%	44.99%	36.77%
+V,-S,-PCA	987.55	50.60%	30.78%	35.13%	44.99%	36.14%
-V,+S,-PCA	194.04	49.49%	44.62%	35.50%	44.99%	35.55%
+V, +S, -PCA	94.45	95.91%	85.05%	98.62%	99.52%	98.88%
+V, +S, +PCA	86.82	99.47%	98.09%	98.51%	99.52%	90.19%

+ = with, - = without; V=Vectorization, S=Scaling, PCA=Principal Component Analysis

As an example, we performed experimentation on LBP features and observed following findings.

- When classifiers are trained with raw observations of features i.e. without vectorization, without scaling, and without PCA, then classifiers do not perform better and consuming large amount of time in the order of 11 times as comparison with vectorization, with scaled features, and with PCA!
- When classifiers are trained with vectorized features but without scaling and without PCA, then also no performance is seen but small decrease in time consumption in the order of 9 times as comparison with vectorization, with scaled features, and with PCA.
- When classifiers are trained with scaled features but without vectorization and without PCA, then time consumption by classifiers are largely decreased but at the same time no performance increase seen.
- When classifiers are trained with vectorization and with scaled features without PCA, large performance improvement is seen and are consuming less time in the order of twice as comparison with vectorization, with scaled features, and with PCA.
- When classifiers are trained with vectorization, with scaled features, and with PCA, then performances of classifiers are greatly increased with very short time consumption.

Therefore, we continued to adopt this VSPCA scheme for all of our experimentation.

A. *Results with Eccentricity Features*

1) Results obtained using eccentricity features without PCA

Experimentation is carried out with eccentricity features and following results obtained. Using vectorized features without and with feature scaling and without PCA, five classifier performance are seen in (TABLE III.).

TABLE III. CLASSIFIER ACCURACY USING ECCENTRICITY FEATURES

All accuracies are in percent.

Accuracy	RFC	SVM	KNN	LogRegr	Dtree
No Scaling	99.52b	99.52	98.78	98.78	99.41
Scaling	99.36	69.26	98.56	99.52	99.52

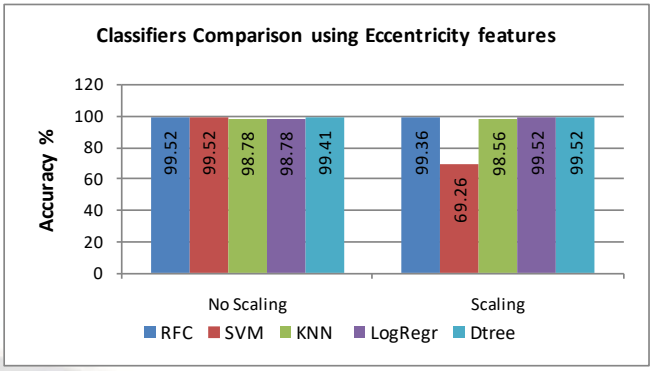

Figure 2. Classifier accuracy with eccentricity features

An accuracy of 99.52% is observed for Logistic Regression (LogRegr) and Decision Tree (Dtree) classifier accuracy using scaled eccentricity features. whereas same is observed for Random Forest Classifier (RFC) and Support Vector Machine (SVM) using non-scaled eccentricity features as in (Fig. 2). K-Nearest Neighbor (KNN) classifier given accuracy of 98.78% in case of non-scaled eccentricity features.

2) Results obtained using eccentricity features with PCA

All observations are obtained using VSPCA scheme.

TABLE IV. ACCURACY OF CLASSIFIERS USING ECCENTRICITY FEATURES WITH PCA

PCA	RFC	SVM	KNN	LogRegr	DTree
2	95.70	72.39	88.87	47.64	98.67
5	99.52	91.20	98.51	95.54	99.36
7	99.52	94.70	98.62	98.67	99.31
9	99.52	98.03	98.62	99.41	99.47
11	99.52	99.52	98.62	99.52	99.52
13	99.52	99.52	98.56	99.52	99.52
15	99.52	99.52	98.56	99.52	99.41
17	99.52	99.52	98.56	99.52	99.52
25	99.52	99.47	98.62	99.52	99.47
30	99.52	99.52	98.56	99.52	99.47
35	99.52	99.41	98.56	99.52	99.52
40	99.52	99.47	98.56	99.52	99.47
45	99.52	99.52	98.56	99.52	99.52
50	99.52	99.52	98.56	99.52	99.47
55	99.52	99.52	98.56	99.52	99.52
60	99.52	99.52	98.56	99.52	99.47
65	99.52	99.41	98.62	99.52	99.52
70	99.52	99.36	98.56	99.52	99.52

It is seen from (TABLE IV.) that, RFC stabilized to 99.52% accuracy with just 5 principal components, whereas SVM, LogRegr, and Dtree stabilized their accuracy to 99.52% at 11 principal components. KNN accuracy did not go beyond 98.62% which is obtained at 7 principal components.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

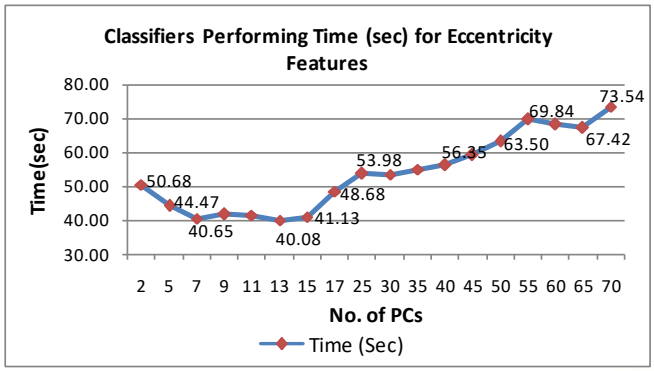

Figure 3. Training time (sec.) for classifiers with varying PCA

It is observed from Fig. 3 and (TABLE IV.) that, while classifiers stabilized their accuracies with PCs ranging between 5 to 11, time consumption by these classifiers is very less and 40.65 seconds lowest time recorded at 7 PC.

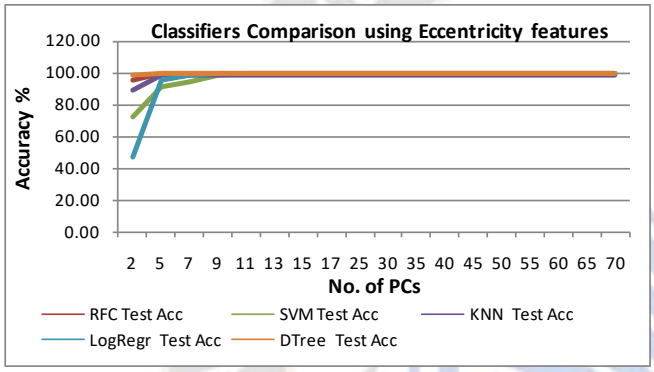

Figure 4. Classifier accuracy using eccentricity features with VSPCA

B. *Results with Euler Number and Eccentricity codes*

1) Results obtained using Euler number and eccentricity features without PCA

Random Forest Classifier (RFC), SVM, KNN, LogRegr, and Dtree these five classifiers are trained with Euler Number and Eccentricity features (Euler-Ecc) together and following results obtained. Using vectorized features without and with feature scaling but without PCA, five classifier performance are seen in (TABLE V.).

TABLE V. CLASSIFIER ACCURACY USING EULER-ECC FEATURES

Accuracy	RFC	SVM	KNN	LogRegr	Dtree
No Scaling	99.20c	88.18	94.59	98.78	99.47
Scaling	99.15	69.26	98.56	99.52	98.88

All accuracies are in percent

Five classifiers accuracy using Euler-Ecc features with and without scaling is observed and all classifiers perform different.

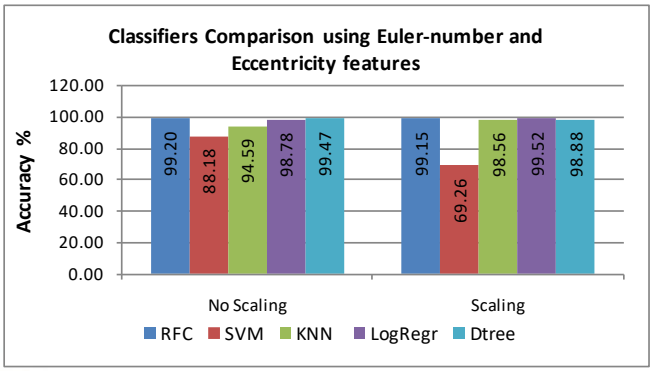

Figure 5. Classifier accuracy comparison using Euler-Ecc features

It is seen from Fig. 5 that, LogRegr perform better which produced 99.52% accuracy in case of scaled Euler-Ecc features than Dtree and RFC which produced 99.47% and 99.20% accuracy respectively in case of features without scaling. KNN produced 98.56% accuracy in case of scaled features whereas SVM produced lowest 88.81% accuracy without scaled Euler-Ecc features.

2) Results obtained using Euler number and eccentricity features with PCA.

All observations are obtained using VSPCA scheme.

TABLE VI. FIVE CLASSIFIER ACCURACY USING EULER-ECC FEATURES WITH VSPCA SCHEME

PCA	RFC	SVM	KNN	LogRegr	Dtree
2	60.30	42.81	54.95	42.02	57.97
5	99.52	92.79	97.45	93.05	98.83
7	99.52	97.61	98.46	98.30	98.99
9	99.52	98.56	98.51	99.25	98.99
11	99.52	99.09	98.62	99.41	99.36
13	99.52	99.25	98.56	99.47	99.31
15	99.52	99.31	98.62	99.47	98.99
17	99.52	99.36	98.56	99.52	99.25
25	99.52	99.09	98.56	99.52	99.15
30	99.52	99.15	99.56	99.52	99.04
35	99.52	99.09	98.56	99.52	99.25
40	99.52	99.20	98.56	99.52	99.25
45	99.52	99.15	98.62	99.52	99.31
50	99.52	99.09	98.56	99.52	99.20
55	99.52	99.20	98.56	99.52	99.36
60	99.52	99.25	98.62	99.52	99.41
65	99.52	99.09	98.56	99.52	99.36
70	99.52	99.31	98.56	99.52	99.47

It is seen from (TABLE VI.) that, using VSPCA scheme, RFC accuracy stabilized to 99.52% with only 5 Principal Components which observed highest among other classifier accuracies on the same Principal Components (PCs). LogRegr stabilized its accuracy to 99.52% at 17 principal components (PCs) which are more PCs than it required for RFC. SVM produced 99.31% accuracy at 15 PCs, Dtree produced 99.47% accuracy at 70 PCs. KNN did not go beyond 98.62% which is obtained at 7 PCs tested through 70 PCs.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

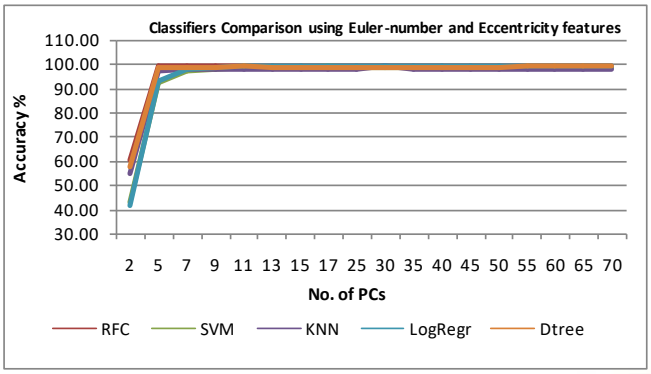

Figure 6. Accuracy score using Euler-Ecc features with VSPCA scheme

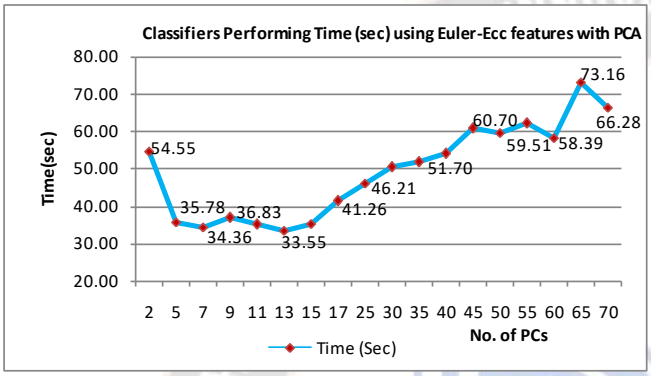

Figure 7. Classifier performing time using Euler-Ecc features with VSPCA scheme

It is observed from Fig. 7 that classifier performing time is less in the range of 5 to 17 principal components while operating on Euler-Ecc features with VSPCA scheme.

C. Results using Horizontal Histogram features

1) Results of Classifiers using Horizontal Histogram features without PCA

Now we will see the results obtained from five classifiers using Horizontal Histogram features without PCA.

TABLE VII. CLASSIFIER ACCURACY USING HORIZONTAL HISTOGRAM FEATURES

Accuracy	RFC	SVM	KNN	LogRegr	Dtree
No Scaling	97.29	56.96	44.62	90.24	98.94
Scaling	96.92	86.85	98.62	99.52	98.94

It is seen from (TABLE VII.) that, LogRegr produced 99.52% accuracy which is highest among other classifiers using Horizontal Histogram features without PCA.

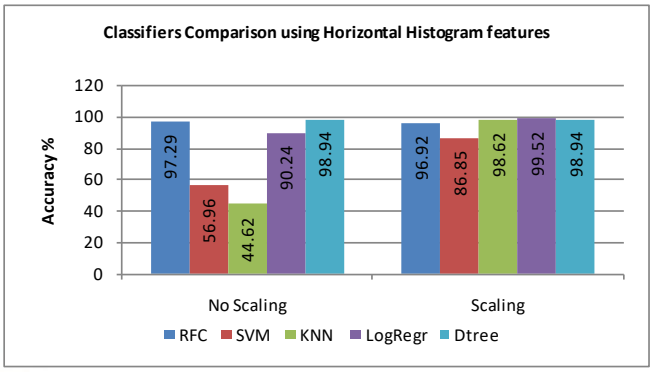

Figure 8. Classifier accuracy using Horizontal Histogram features

2) Results of Classifiers using Horizontal Histogram features with PCA

All results are obtained with VSPCA scheme.

TABLE VIII. CLASSIFIER ACCURACY WITH HORIZONTAL HISTOGRAM FEATURES WITH VSPCA SCHEME

PCA	RFC	SVM	KNN	LogRegr	Dtree
2	31.42	20.24	19.23	20.29	29.19
5	55.85	46.42	46.79	40.11	46.74
7	65.44	58.71	59.98	51.13	51.93
9	83.73	78.16	77.15	74.29	70.21
11	98.25	90.14	87.81	92.31	94.8
13	99.36	94.27	93.21	96.87	98.14
15	99.52	95.49	96.07	98.83	98.25
17	99.52	96.66	96.71	98.72	98.19
25	99.52	97.66	98.03	99.36	98.09
30	99.52	97.13	98.30	99.52	98.88
35	99.52	97.45	98.41	99.41	98.78
40	99.52	97.08	98.56	99.52	99.31
45	99.52	97.13	98.46	99.52	98.72
50	99.52	97.35	98.51	99.52	99.31
55	99.52	96.71	98.56	99.52	98.67
60	99.52	96.92	98.67	99.52	98.94
65	99.52	96.71	98.62	99.52	99.09
70	99.52	96.55	98.62	99.52	98.94

It is seen from (TABLE VIII.) that, using VSPCA scheme, RFC accuracy stabilized to 99.52% with 15 principal components. LogRegr accuracy stabilized to 99.52% from 40 principal components which consumed more number of principal components than RFC to stabilize. Accuracy obtained from SVM is found oscillating between 96 and 98% from 17 principal components. In the same way, accuracy of Dtree is seen oscillating between 98 and 99% from 13 principal components. Accuracy of KNN is observed not uplifted above 98.62% consuming all 70 principal components.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

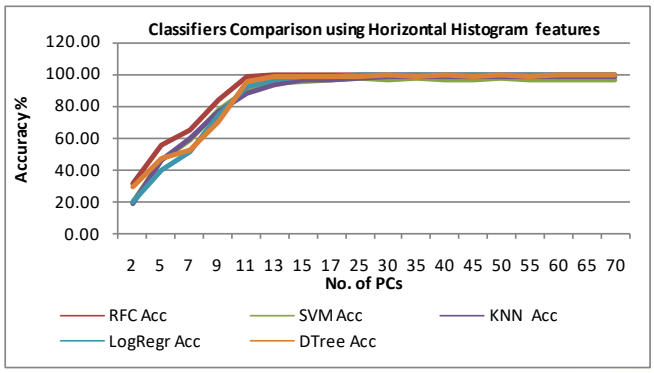

Figure 9. Classifier accuracy using Horizontal Histogram features with VSPCA scheme

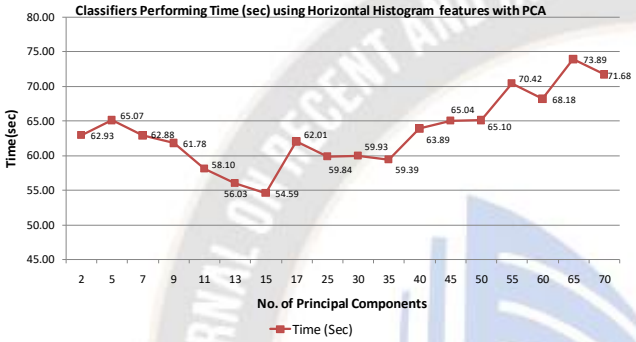

Figure 10. Classifier performing time using Horizontal Histogram features with VSPCA scheme

It is seen from classifier performing time given in Fig. 10 that, the time is ranging between 11 to 17 principal components during which it is less.

D. Results using Vertical Histogram features

1) Results of Classifiers using Vertical Histogram features without PCA

TABLE IX. CLASSIFIER ACCURACY USING VERTICAL HISTOGRAM FEATURES

Accuracy	RFC	SVM	KNN	LogRegr	Dtree
No Scaling	96.55	51.45	44.35	90.62	98.94
Scaling	96.18	91.52	98.62	99.52	98.88

From (TABLE IX.) it is seen that, LogRegr performs better and produced 99.52% accuracy with scaled Vertical Histogram features, below that Dtree classifier produced 98.94% accuracy without scaled Vertical Histogram features. Classifier KNN performed somewhat below Dtree and produced 98.62% accuracy with scaled features. Classifier RFC produced 96.55% accuracy without scaled features where SVM performed lowest and produced 91.52% accuracy with scaled features.

When we see classifier performing time, 124.31 seconds time required in case of without scaled features whereas 89.64 seconds are required in case of scaled features.

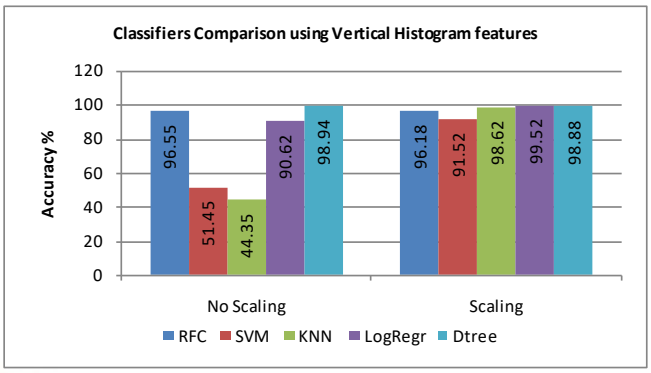

Figure 11. Classifier accuracy using Vertical Histogram features without PCA

Further, it can be seen from Fig. 11 that, majority of classifiers performed well in vectorized scaled features than that of non-scaled Vertical Histogram features.

2) Results of Classifiers using Vertical Histogram features with PCA

Experimentation is carried out using RFC, SVM, KNN LogRegr, Dtree classifiers over Vertical Histogram features using VSPCA scheme. The results obtained are with VSPCA scheme.

TABLE X. CLASSIFIER ACCURACY WITH VERTICAL HISTOGRAM FEATURES WITH VSPCA SCHEME

PCA	RFC	SVM	KNN	LogRegr	Dtree
2	35.55	23.26	23.42	23.95	33.59
5	58.61	48.01	48.27	44.35	48.38
7	67.30	58.71	59.03	54.10	54.84
9	94.43	80.49	78.69	83.09	87.22
11	99.15	86.75	86.96	92.52	97.61
13	99.52	92.63	93.85	97.19	98.25
15	99.52	94.54	95.49	98.46	98.19
17	99.52	96.34	96.44	98.67	98.83
25	99.52	94.96	97.88	99.31	98.56
30	99.52	96.13	97.82	99.41	99.04
35	99.52	95.70	97.98	99.52	99.04
40	99.52	96.82	98.19	99.47	98.62
45	99.52	96.29	98.30	99.52	99.31
50	99.52	95.44	98.25	99.52	99.31
55	99.52	95.07	98.30	99.52	99.36
60	99.52	95.01	98.35	99.52	98.72
65	99.52	95.28	98.30	99.52	98.72
70	99.52	95.81	98.30	99.52	99.15

It is seen from (TABLE X.) that, using VSPCA scheme, RFC classifier provided 99.52% accuracy which is stabilized from 13 principal components. Accuracy obtained from LogRegr classifier stabilized to 99.52% from 45 principal components which shows that LogRegr classifier required more number of principal components than RFC to stabilize. Dtree classifier accuracy found 99.36% obtained at 55 principal components, whereas KNN produced it 98.30% at 45 PCs. SVM is found oscillating between 95% and 97% from 15 principal components.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

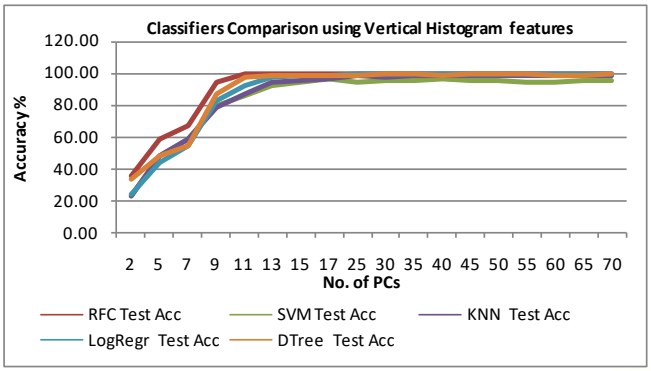

Figure 12. Classifier accuracy using Vertical Histogram features using PCA

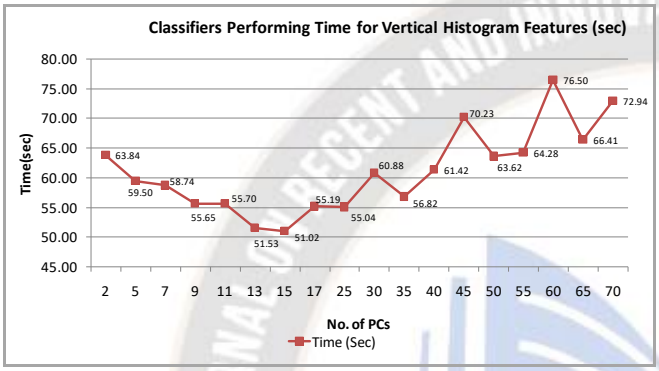

Figure 13. Classifier performing time using Vertical Histogram features using PCA

It is seen from Fig. 13 that, performing time to classifier is seen low within the range of 11 to 17 principal components.

E. Results using HOG features

1) Results of Classifiers using Histograms of Oriented Gradient (HOG) features without PCA

TABLE XI. CLASSIFIER ACCURACY USING HOG FEATURES WITHOUT PCA

PCA	RFC	SVM	KNN	LogRegr	Dtree
No Scaling	94.22	94.75	76.89	97.61	99.25
Scaling	93.79	95.65	97.93	99.52	99.09

From (TABLE XI.) it is seen that, LogRegr performs better and produced 99.52% accuracy with scaled HOG features, below that Dtree classifier produced 99.25% accuracy without feature scaling. Classifier KNN performed somewhat below Dtree and produced 97.93% accuracy with scaled features. Classifier SVM produced 95.65% accuracy with scaled features where RFC performed lowest and produced 94.22% accuracy without scaled HOG features.

Classifier performing time, 191.20 seconds time required in case of without scaled features whereas 166.45 seconds are required in case of scaled HOG features. Hence, it is observed that classifier performing time is less required in case of scaled features than without scaled features.

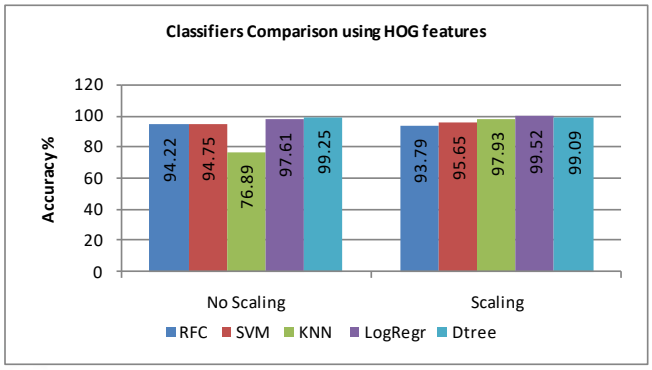

Figure 14. Classifier accuracy using HOG features without PCA

2) Results of Classifiers using Histograms of Oriented Gradient (HOG) features with PCA

Experimentation is performed on HOG features using RFC, SVM, KNN LogRegr, Dtree classifiers with VSPCA scheme and results are obtained as below.

TABLE XII. CLASSIFIER ACCURACY USING HOG FEATURES WITH PCA

PCA	RFC	SVM	KNN	LogRegr	Dtree
2	33.33	19.39	21.19	20.66	31.10
5	53.41	40.54	42.97	37.04	44.83
7	59.45	49.28	49.65	47.16	48.86
9	65.02	55.37	54.95	55.00	52.19
11	68.20	61.15	59.77	61.10	53.73
13	70.69	64.81	63.38	65.60	54.58
15	73.82	69.84	66.50	72.28	56.43
17	78.06	76.31	71.59	79.01	59.61
25	95.49	92.05	85.95	96.71	75.09
30	98.88	95.86	91.94	98.30	84.26
35	99.36	97.45	94.22	99.20	89.71
40	99.41	97.77	95.44	99.25	91.67
45	99.52	97.88	95.44	99.41	94.01
50	99.52	97.56	95.81	99.36	97.19
55	99.52	97.66	95.91	99.47	96.34
60	99.52	97.45	95.86	99.47	95.76
65	99.52	97.29	96.02	99.52	97.03
70	99.52	97.19	96.13	99.52	97.61

It is seen from (TABLE XII.) that, using VSPCA scheme, RFC classifier provided 99.52% accuracy which is stabilized from 45 principal components. Accuracy obtained from LogRegr classifier stabilized to 99.52% from 65 principal components which shows that LogRegr classifier required more number of principal components than RFC to stabilize. Dtree classifier accuracy found 97.61% obtained at 70 principal components, whereas SVM classifier produced it 97.77% at 40 PCs. Accuracy from KNN is observed 96.13% at 70 principal components.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

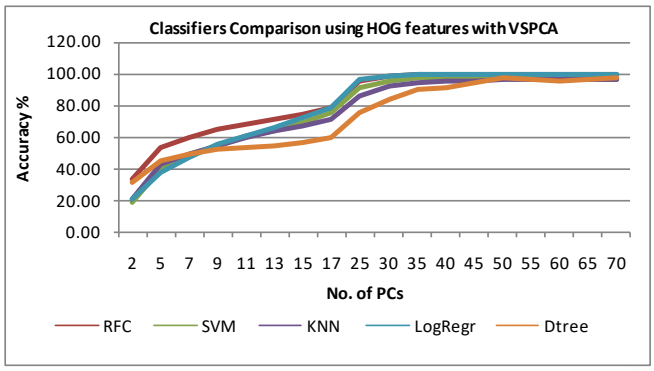

Figure 15. Classifier accuracy using HOG features with VSPCA scheme

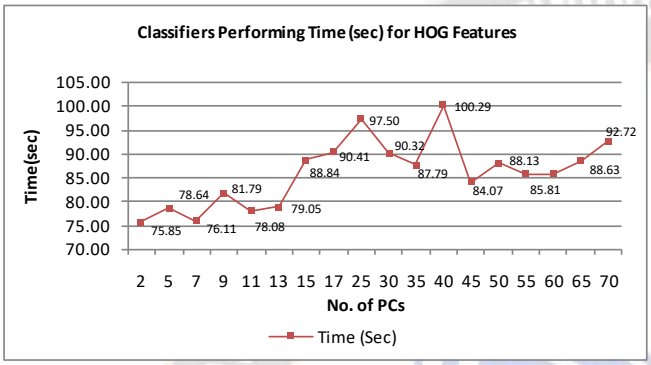

Figure 16. Classifier performing time using HOG features with VSPCA scheme

F. *Results using LBP features*

1) *Results of Classifiers using LBP features without PCA*

TABLE XIII. CLASSIFIER ACCURACY USING LBP FEATURES WITHOUT PCA

Accuracy	RFC	SVM	KNN	LogRegr	Dtree
No Scaling	96.39	30.89	35.98	39.79	98.94
Scaling	96.39	85.05	98.62	99.52	98.78

From (TABLE XIII.) it is observed that, LogRegr performs better and produced 99.52% accuracy with scaled LBP features which is jumped from 39.79% accuracy obtained using non-scaled features. Below that, Dtree classifier produced 98.94% accuracy without scaled feature and observed not much variation in accuracy obtained from scaled features. Classifier KNN performed somewhat below Dtree and produced 98.62% accuracy with scaled features but improved from 35.98% accuracy using non scaled features. Classifier RFC produced 96.39% accuracy both with and without scaled features. SVM found performed lowest in the same scenario and produced 85.05% accuracy with scaled LBP features which is seen improved from using non scaled features from 30.89% accuracy.

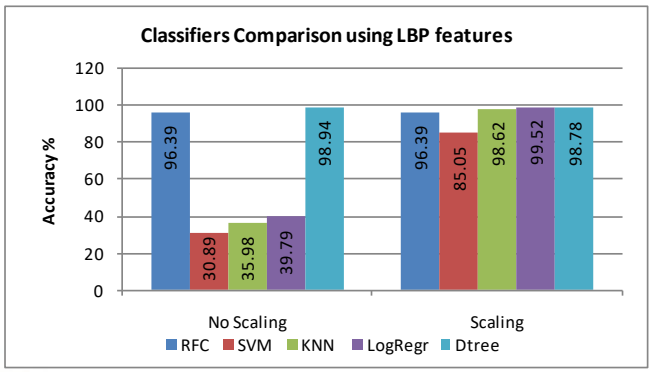

Figure 17. Classifier accuracy using LBP features without PCA

It is seen from Fig. 17 that, significant improvement seen in accuracies obtained from classifiers SVM, KNN, and LogRegr in case of scaled LBP features, whereas, RFC and Dtree classifier accuracies found stabilized in both scaled and without scaled LBP features.

It is also observed that, classifier performing time is 114.38 seconds in case of non scaled features and 105.08 seconds in case of scaled features which are found reduced when using scaled features.

2) *Results of Classifiers using LBP features with PCA*

Experimentation is performed on LBP features using RFC, SVM, KNN LogRegr, Dtree classifiers with VSPCA scheme and results are obtained as below.

TABLE XIV. CLASSIFIER ACCURACY USING LBP FEATURES WITH PCA

PCA	RFC	SVM	KNN	LogRegr	DTree
2	31.21	19.66	18.22	21.19	29.99
5	54.68	45.99	41.86	45.99	47.00
7	64.81	57.71	54.21	56.70	53.84
9	72.28	66.87	64.28	66.61	58.02
11	77.90	76.09	72.54	77.84	61.95
13	82.61	83.46	80.12	86.85	66.50
15	87.06	87.65	86.53	91.04	70.05
17	90.51	92.36	91.09	94.70	72.07
25	97.50	96.60	96.55	98.46	75.56
30	98.41	97.08	97.13	98.94	79.22
35	99.25	98.35	98.03	99.25	82.03
40	99.36	98.41	98.09	99.36	84.73
45	99.31	97.93	98.35	99.36	86.32
50	99.41	98.25	98.35	99.47	85.21
55	99.47	98.09	98.14	99.47	89.29
60	99.41	98.25	98.46	99.41	89.13
65	99.47	98.56	98.41	99.52	90.03
70	99.47	97.93	98.51	99.47	90.19
75	99.52	98.03	98.46	99.52	93.32

It is seen from (TABLE XIV.) that, using VSPCA scheme, RFC classifier provided 99.52% accuracy which is stabilized from 75 principal components. Accuracy obtained from LogRegr classifier stabilized to 99.52% from 65 principal components. SVM classifier accuracy found 98.56% obtained

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

at 65 principal components, whereas KNN classifier produced it 98.46% at 60 PCs. Accuracy from Dtree is observed 93.32% at 75 principal components.

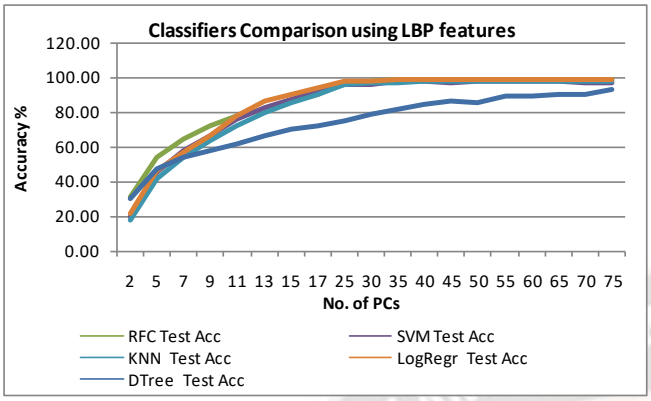

Figure 18. Classifier accuracy using LBP features with PCA

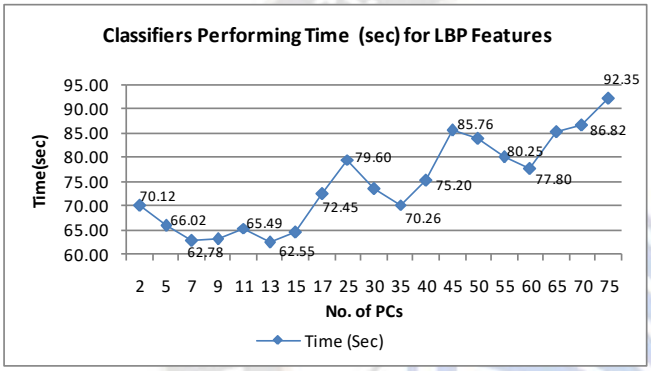

Figure 19. Classifier performing time using LBP features with PCA

G. Results using Statistical features

1) Results of Classifiers using Statistical features without PCA

Statistical features are experimented using RFC, SVM, KNN LogRegr, Dtree classifiers without PCA scheme and results are obtained as below.

TABLE XV. CLASSIFIER ACCURACY USING STATISTICAL FEATURES WITHOUT PCA

PCA	RFC	SVM	KNN	LogRegr	Dtree
No Scaling	93.16	29.57	32.85	25.11	98.78
Scaling	93.26	96.44	97.61	99.52	98.78

From (TABLE XV.) it is observed that, all five classifiers RFC, SVM, KNN, LogRegr and Dtree are performing well when provided scaled Statistical features. Very high jump of accuracy is seen for SVM, KNN, and LogRegr than RFC and Dtree where these two classifiers are less affected by scaled Statistical features without applying PCA. Classifier accuracies for RFC, SVM, KNN, LogRegr and Dtree are 93.26%, 96.44%, 97.61%, 99.52% and 98.78% respectively. Accuracy from LogRegr classifier is seen highest among accuracy from other classifiers.

It is also observed that, classifier performing time is 132.66 seconds in case of non scaled features and 93.78 seconds in case of scaled statistical features which are found reduced when using scaled features.

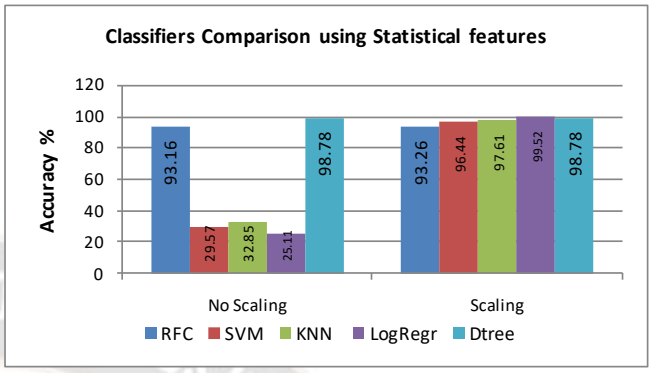

Figure 20. Classifier accuracy using Statistical features without PCA

2) Results of Classifiers using Statistical features with PCA

Experimentation on LBP features are carried out using RFC, SVM, KNN LogRegr, Dtree classifiers with VSPCA scheme and results are obtained as below.

TABLE XVI. CLASSIFIER ACCURACY USING STATISTICAL FEATURES WITH PCA

PCA	RFC	SVM	KNN	LogRegr	Dtree
2	32.90	24.27	21.19	24.58	30.36
5	54.79	45.25	44.14	42.39	45.78
7	64.01	57.07	54.26	52.88	51.56
9	68.99	64.12	61.79	60.09	53.94
11	71.96	68.04	65.23	64.54	56.06
13	75.14	73.55	70.48	70.80	59.40
15	77.26	77.21	73.18	75.19	60.25
17	77.95	79.06	74.29	78.69	59.08
25	93.21	88.97	82.82	94.85	73.02
30	98.88	93.05	87.33	97.66	87.12
35	99.25	94.11	90.35	98.51	93.58
40	99.47	95.17	92.63	98.88	95.76
45	99.52	95.38	93.48	98.94	96.60
50	99.52	96.02	94.22	99.20	96.76
55	99.52	95.97	94.54	99.15	96.87
60	99.52	95.70	94.70	99.31	97.40
65	99.52	95.70	95.65	99.36	97.29
70	99.52	96.07	95.86	99.25	97.93

It is seen from (TABLE XVI.) that, using VSPCA scheme, RFC classifier provided 99.52% accuracy which is stabilized from 45 principal components. Accuracy obtained from LogRegr classifier is found 99.36% at 65 principal components. Accuracies from Dtree, SVM, and KNN classifier are found 97.93%, 96.07% and 95.86% found at 70 principal components.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

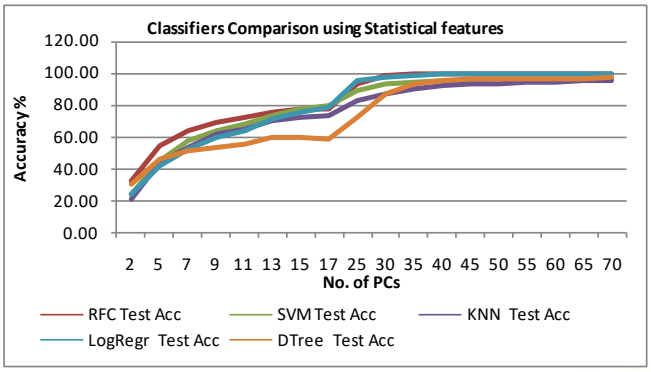

Figure 21. Classifier accuracy using Statistical features with PCA

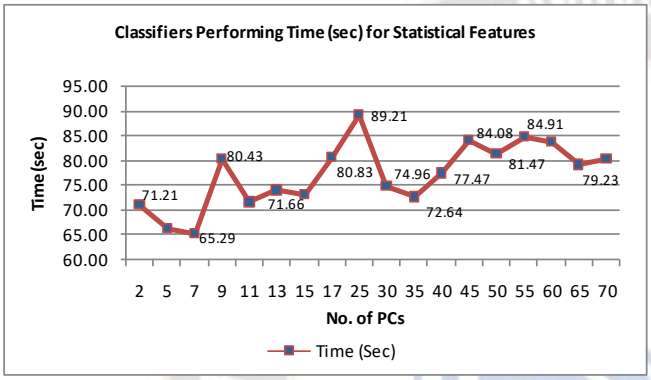

Figure 22. Classifier performing time (sec) for Statistical features with PCA

H. Overall Results Discussion

TABLE XVII. SINGLE CLASSIFIER APPROACH RESULTS WITH ALL FEATURES

Method vs feature	No Scaling	Scaling	VSPCA Scheme
Eccentricity	99.52 (RFC), 99.52(SVM), 99.41 (Dtree), 98.78 (KNN), 98.78 (LogRegr)	99.52 (LogRegr) 99.52 (Dtree), 99.36 (RFC), 98.56 (KNN), 69.26 (SVM)	99.52 (RFC-5PC), 99.52 (SVM-11PC), 99.52 (LogRegr-11PC), 99.52 (Dtree-11PC) 98.62 (KNN-7PC)
Euler number and Eccentricity	99.47 (Dtree), 99.2 (RFC), 98.78 (LogRegr), 94.59 (KNN), 88.18 (SVM)	99.52 (logRegr), 99.15 (RFC), 98.88 (Dtree), 98.56 (KNN), 69.26 (SVM)	99.52 (RFC-5PC), 99.52 (LogRegr-17PC), 99.47 (Dtree-70PC), 99.31 (SVM-15PC), 98.62 (KNN-11PC)
LBP	98.94 (Dtree), 96.39 (RFC), 39.75 (LogRegr), 35.98 (KNN), 30.89 (SVM)	99.52 (LogRegr), 98.78 (Dtree), 98.62 (KNN), 96.39 (RFC), 85.05 (SVM)	99.52 (LogRegr-65PC), 99.52 (RFC-75PC), 98.51 (KNN-70PC), 98.56 (SVM-65PC), 90.19 (Dtree-70PC)
HOG	99.25 (DTree), 97.61 (LogRegr), 94.22 (RFC), 94.75 (SVM), 76.89 (KNN)	99.52 (LogRegr), 99.09 (DTree), 97.93 (KNN), 95.65 (SVM), 93.79 (RFC)	99.52 (RFC-45PC), 99.52 (LogRegr-65PC), 97.88 (SVM-45PC), 97.61 (DTree-70PC) 96.13 (KNN-70PC)
Horizontal Histogram	98.94 (DTree), 97.29 (RFC), 90.24 (LogRegr), 56.96 (SVM), 44.62 (KNN)	99.52 (LogRegr), 98.94 (DTree), 98.62 (KNN), 96.92 (RFC), 86.85 (SVM)	99.52 (RFC-15PC), 99.52 (LogRegr-30PC), 99.31 (DTree-40PC), 98.67 (KNN-60PC), 97.66 (SVM-25PC)
Vertical Histogram	98.94 (DTree), 96.55 (RFC), 90.62 (LogRegr), 51.45 (SVM), 44.35 (KNN)	99.52 (LogRegr), 98.88 (DTree), 98.62 (KNN), 96.18 (RFC), 94.52 (SVM)	99.52 (RFC-13PC), 99.52 (LogRegr-45PC), 99.36 (DTree-55PC), 98.35 (KNN-60PC), 96.82 (SVM-40PC)
Statistical features	98.78 (DTree), 93.16 (RFC), 32.85 (KNN), 29.57 (SVM), 25.11 (LogRegr)	99.52 (LogRegr), 98.78 (DTree), 97.61 (KNN), 96.44 (SVM), 92.79 (RFC)	99.52 (RFC-45PC), 99.36 (logRegr-65PC), 97.93 (DTree-70PC), 96.07 (SVM-70PC), 95.86 (KNN-70PC)

As indicated in (TABLE XVII.), it is seen that by using without scaled features, very few classifiers reached to 99.52% accuracy in all the feature sets. However, when features are scaled, at least one classifier produced 99.52% accuracy within each feature set. But when features are operated with VSPCA scheme, majority of classifiers in each feature group produced higher accuracy of 99.52% or near. Therefore, VSPCA scheme outperforms than no scaling and scaling feature scheme.

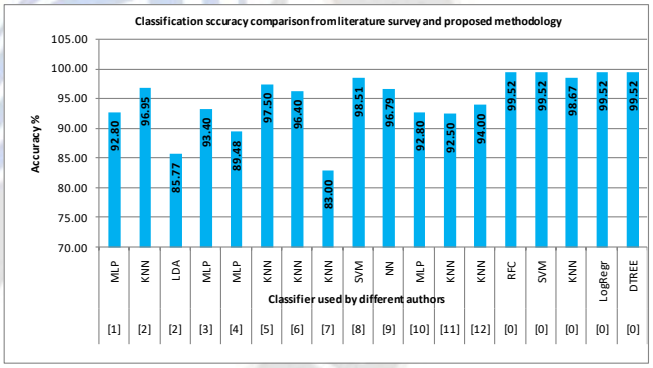

Figure 23. Classifier accuracy comparison of proposed method [0] with different authors [1]-[12].

The work carried out by other authors [1]-[12] given in Fig. 23 shows that, our VSPCA scheme worked better than their accuracies and obtained using proposed method is found 99.52% which are shown as last five bars indicated [0] as our proposed work. The work carried out by [8] obtained 98.51% accuracy which is lower than our obtained accuracy in case of single classifier approach.

CONCLUSION

As we used Five different classifiers such as Random Forest Classifier (RFC), Support Vector Machine (SVM), K Nearest Neighbor Classifier (KNN), Logistic Regression Classifier (LogRegr), Decision Tree Classifier (DTree) with Seven different feature sets namely Eccentricity features, Euler-Ecc features, Horizontal Histogram features, Vertical Histogram features, HOG Features, LBP features, and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

Statistical features along over three different methods such as without scaling, scaling and PCA, we have clearly detailed performance of each of these methods precisely considering classifier and feature set combinations along with vectorization.

Performance in Non-Scaling method varies from 25.11 (LogRegr) to 99.52% (RFC, SVM). Performance in Scaling method varies from 69.26% (SVM) to 99.52% (LogRegr, Dtree). Performance in proposed VSPCA method stands at 99.52% for many classifiers except for KNN which shows 98.67% accuracy. In proposed VSPCA method, classifier accuracies for most of the feature set combinations stands at 99.52% which is observed higher than that of non-scaling or scaling methods and also observed higher from earlier work carried out by different researchers [1]-[12] as depicted in Fig. 23.

Beauty of this research is in comparison of lowest classifier from non-scaling method with VSPCA method. In LBP feature set, SVM shows 30.89% of accuracy for non scaling method whereas the same SVM shows 98.56% in VSPCA method with 65 PCs. In Horizontal Histogram feature set, KNN shows 44.62% of accuracy for non-scaling method whereas it shows 98.67% in VSPCA method with 60 PCs. In vertical histogram feature set, KNN shows 44.35% of accuracy whereas it shows 98.35% in VSPCA method with 60 PCs. When Logistic Regression classifier applied with Statistical features it shows poor performance giving an accuracy of 25.11% against non-scaling method. Further it shows 99.36% accuracy in VSPCA method. Also it is observed that the accuracy is improved to 99.52% for Eccentricity, Euler number and Eccentricity, LBP, HOG, Horizontal Histogram, and Vertical Histogram features.

It is concluded based on experimentation and classifiers considered that 99.52% accuracy achieved using VSPCA scheme.

REFERENCES

[1] Sk Md Obaidullah, Supratik Kundu Das, Kaushik Roy, "A System for Handwritten Script Identification from Indian Document", Researchgate: Journal of Pattern Recognition Research 8 (2013) 1-12, Researchgate, 2013, pp.1-12.

[2] Mallikarjun Hangarge, Santosh K.C., Rajmohan Pardeshi, "Directional Discrete Cosine Transform for Handwritten Script Identification", 12th International Conference on Document Analysis and Recognition (ICDAR), 2013, IEEE, 2013, pp.344-348.

[3] Deepti Khanduja, Neeta Nain, and Subhash Panwar, "A Hybrid Feature Extraction Algorithm for Devanagari Script", ACM Trans. Asian Low-Resour. Lang. Inf. Process., Vol. 15, No. 1, Article 2, ACM, 2015, pp.1-10.

[4] K. Roy, S. Kundu Das, Sk Md Obaidullah, "Script Identification from Handwritten Document", Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, 2011, Researchgate, 2011, pp.1-5.

[5] Hiremath P. S., Shivashankar S., Jagdeesh D. Pujari, V. Mouneswara, "Script identification in a handwritten document image using texture features", 2010 IEEE 2nd International Advance Computing Conference, IEEE Explore, IEEE, 2010, pp.110-114.

[6] Dr. Antino Marelino. (2014). Customer Satisfaction Analysis based on Customer Relationship Management. International Journal of New Practices in Management and Engineering, 3(01), 07 - 12. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/26

[7] G. G. Rajput, Anita H. B, "Handwritten Script Recognition using DCT and Wavelet Features at Block Level", RTIPPR 2010, IJCA - ResearchGate, 2010, pp.158-163.

[8] Mallikarjun Hangarge, B.V.Dhandra, "Offline handwritten script identification in document images", IJCA 2010 Vol.4,No.6, IJCA, 2010, pp.6-10.

[9] Sukalpa Chanda, Srikanta Pal, Katrin Franke, Umapada Pal, "Two-stage Approach for Word-wise Script Identification", 2009 10th ICDAR, IEEE, 2009, pp.926-930.

[10] Akram Baig, M. M. . (2023). An Evaluation of Major Fault Tolerance Techniques Used on High Performance Computing (HPC) Applications. International Journal of Intelligent Systems and Applications in Engineering, 11(3s), 320–328. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2696

[11] K. Roy, K. Majumder, "Trilingual Script Separation of Handwritten Postal Document", Sixth Indian Conference on Computer Vision, Graphics & Image Processing, 2008. ICVGIP '08. , IEEE, 2008, pp.693-700.

[12] Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, Dipak Kumar Basu, Mahantapas Kundu, "Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition", 2008 IEEE Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA December 8-10., IEEE, 2008, pp.342-1-342-6.

[13] B.V.Dhandra, Mallikarjun Hangarge, "Global and Local Features Based Handwritten Text Words and Numerals Script Identification", International Conference on Computational Intelligence and Multimedia Applications 2007, IEEE, 2007, pp.471-475.

[14] Gopal Datt Joshi, Saurabh Garg, Jayanthi Sivaswamy, "A generalised framework for script identification", IJDAR (2007), Springer, 2007, pp.55-68.

[15] More Vijay, Kharat M U, Gumaste S V, "Study of different features and classification techniques for recognition of handwritten Devanagari text", International Journal of Engineering & Technology (IJET-UAE), ISSN-1055-1059, Vol 7, No. 4, Issue 19, DOI:10.14419/ijet.v7i4.19.28285, 2018, pp.1055-1059.

[16] Vijay More, Madan Kharat, Shyamrao Gumaste, "Segmentation Of Devanagari Handwritten Text Using Thresholding Approach", International Journal of Scientific & Technology Research (IJSTR), IJSTR, 2020, pp.6594-6605.

[17] Basaligheh, P. (2021). A Novel Multi-Class Technique for Suicide Detection in Twitter Dataset. Machine Learning Applications in Engineering Education and Management, 1(2), 13–20. Retrieved from http://yashikajournals.com/index.php/mlaeem/article/view/14

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8066
***Article Received:** 21 June 2023 **Revised:** 12 August 2023 **Accepted:** 30 August 2023*

[18] Vijay More, Madan Kharat, "Segmentation of Lines and Words of Handwritten Devanagari Text using Connected Components with Statistics Method", Journal of Scientific Research, Institute of Science, Banaras Hindu University, Varanashi, UP, JSR, 2022, DOI:10.37398/JSR.2022.660224, pp.179-188.

[19] Sk Md Obaidullah, Chayan Halder, K. C. Santosh, Nibaran Das, Kaushik Roy, "PHDIndic_11:Page-level handwritten document image dataset of 11 official Indic scripts for script identification", in Multimedia Tools and Applications (MTAP), Springer, Volume 77, Issue 2, 2017, doi:10.1007/s11042-017-4373-y, pp.1643–1678.

[20] Juan Garcia, Guðmundsdóttir Anna, Johansson Anna, Maria Jansen, Anna Wagner. Optimal Decision Making in Supply Chain Management using Machine Learning. Kuwait Journal of Machine Learning, 2(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/209

[21] Gholamy, Afshin; Kreinovich, Vladik; and Kosheleva, Olga, "Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation" (2018). Departmental Technical Reports (CS). 1209. https://scholarworks.utep.edu/cs_techrep/1209.

[22] Heutte, Laurent; Paquet, Thierry; Moreau, Jean Vincent; Lecourtier, Yves; Olivier, Christian, "A structural / statistical feature based vector for handwritten character recognition" Pattern Recognition Letters 19, 1998, Elsevir, pp.629–641.

[23] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 2, pp. 153-158, Feb. 1997, doi: 10.1109/34.574797.