"From pen to pixels: The future of recognition"

- Anonymous

Chapter 4

OFFLINE HANDWRITTEN DEVANAGARI WORD RECOGNITION SYSTEM BASED ON GRADIENT AND STRUCTURAL FEATURES

4.1 INTRODUCTION

Segmenting words into individual characters can be a challenging and time-consuming task when attempting to recognize a word. The segmentation process is plagued by various issues, including overlapping/touching characters, difficulties in identifying the correct segmentation point (segmentation ambiguity) and the presence of cursive handwriting. These issues often result in improper classification of words, leading to poor accuracy in recognition. To overcome these segmentation challenges, the present study adopts a holistic approach to recognizing offline handwritten Devanagari words, thereby bypassing the need for explicit character segmentation. This approach aims to mitigate the aforementioned segmentation issues and improve the overall accuracy of word recognition.

In this chapter, a method is presented for recognizing and identifying offline handwritten Devanagari words. The proposed approach has diverse applications in pattern recognition, such as cheque reading, airline ticket readers, bill processing systems, handwritten address interpretation, signboard translation, and postal automation. The framework follows a holistic approach for the identification and classification of Devanagari handwritten words. Instead of relying on segmentation, the method treats the entire word as a single entity for recognition. The handwritten word images are processed to extract gradient and structural features, utilizing contour-directional histograms. These features serve as important inputs for the recognition process, enabling accurate identification of the Devanagari words without the need for explicit segmentation. In the recognition tasks, three different classifiers, namely Support Vector Machine (SVM), Naive Bayes (NB), and eXtreme Gradient Boosting (XGBoost), are employed. Additionally, experiments are conducted using combined feature vectors obtained from gradient and structural features as input to the classifiers. The proposed framework is evaluated using a corpus consisting of 20,000 words, representing 50 different town names written in the Devanagari script. The performance of the system is assessed based on metrics such as Recognition Accuracy (RA), False Acceptance Rate (FAR), False Rejection Rate (FRR) and Precision (PR). The experimental results indicate that the combination of feature vectors derived from gradient and structural features, in conjunction with the XGBoost classifier, outperforms the use of individual features alone.

Section 4.2 of this chapter outlines the feature extraction techniques employed for recognizing handwritten Devanagari words. In Section 4.3, the various classification techniques used in this study are discussed. The experimental results and corresponding discussions are presented in Section 4.4, providing an evaluation of the performance of the handwritten word recognition system. Section 4.5 briefly outlines the analysis based on the experimental results. Finally, Section 4.6 summarizes the entire chapter, offering a concise overview of the key points discussed.

4.2 FEATURE EXTRACTION TECHNIQUES

The primary objective of feature extraction techniques is to extract relevant and meaningful information from input handwritten images, which can be effectively utilized for classification or recognition purposes. Developing a handwritten word recognition system poses various challenges and issues, as outlined in previous research (Ramachandrula et al., 2012; Singh and Garg, 2021). Among these challenges, feature selection plays a vital role as it significantly impacts the recognition accuracy of the system. Therefore, it is crucial to extract meaningful and significant features from handwritten word images. Extensive literature research has highlighted that the directional and structural information of characters or words can serve as valuable feature vectors for subsequent classification tasks. In this study, gradient (directional) and structural features are specifically considered for recognition purposes, taking advantage of their potential in enhancing the accuracy of handwritten word recognition systems.

4.2.1 Gradient-based Features

The gradient is a vector quantity that comprises both magnitude and directional components. These components can be determined through differentiation along the horizontal and vertical directions (Jindal and Kumar, 2017). The image gradient can be computed using various operators, such as the Sobel, Robertz, and Prewitt operators. In this study, the gradient vector $[G_u, G_v]^T$ is calculated using the Sobel operator, where G_u and G_v represent the horizontal and vertical components of the gradient, respectively. To extract distinctive information from the handwritten word image, G_u and G_v are obtained using two Sobel operator templates (horizontal and vertical) with dimensions of (3×3), as illustrated in Fig. 4.1.

1	2	1
0	0	0
-1	-2	-1

(a)

1	0	-1
2	0	-2
1	0	-1

(b)

Figure 4.1: Sobel masks: **(a)** horizontal and **(b)** vertical components

To process a given handwritten word image (H) of size $M \times N$, an 8-pixel neighborhood around each pixel (i, j) is considered, where i ranges from 1 to M and j ranges from 1 to N. This neighborhood is illustrated in Fig. 4.2.

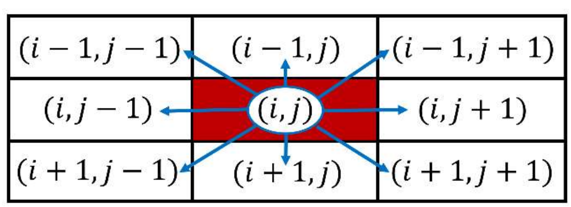


Figure 4.2: An 8-pixel neighborhood at each pixel (i, j)

The Sobel operator templates mentioned earlier are convolved with the pixel values within this neighborhood to calculate $G_u(u, v)$ and $G_v(u, v)$. Mathematically, this can be expressed as given in Eqs. 4.1 and 4.2, respectively:

$$G_u(u,v) = H(i-1,j-1) + 2*H(i-1,j) + H(i-1,j+1) \\ - H(i+1,j-1) - 2*H(i+1,j) - H(i+1,j+1) \tag{4.1}$$

$$G_v(u,v) = H(i-1,j-1) + 2*H(i,j-1) + H(i+1,j-1) \\ - H(i-1,j+1) - 2*H(i,j+1) - H(i+1,j+1) \tag{4.2}$$

Additionally, the magnitude (Refer Eq. 4.3) and direction (Refer Eq. 4.4) of the gradient are computed using the following formulas:

$$|G(i,j)| = \sqrt{[G_u(i,j)]^2 + [G_v(i,j)]^2} \tag{4.3}$$

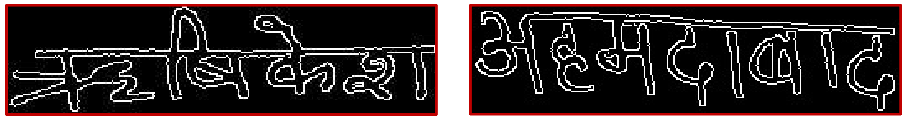
$$\emptyset(i,j) = \tan^{-1}\left[\frac{G_v(i,j)}{G_u(i,j)}\right] \tag{4.4}$$

In this work, the gradient feature vector is generated by extracting the gradient information in a similar manner as implemented by (Jindal and Kumar, 2017) for the Gurumukhi script.

4.2.2 Structural-based Features

Structural features capture the geometric and topological characteristics of handwritten word images by considering their local and global properties (Jindal and Kumar, 2017). The nature of the pattern being classified determines the specific structural features that can be utilized (Jayadevan et al., 2011). In the context of handwritten word recognition tasks, various features can be extracted, including directional histogram-based features, character/word geometry, intersection of line segments/loops, horizontal/vertical projection profiles, structural primitives, left/right profiles, and measurement of cavity features based on jumps/discontinuities in characters or words (Ghosh et al., 2010).

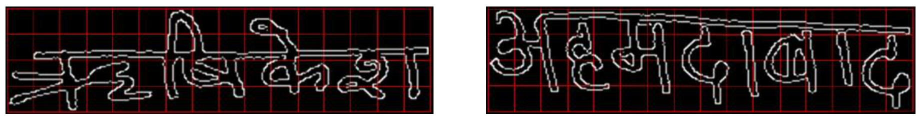
These features directly relate to the shape of the character/word for recognition purposes. In this study, contour-directional histogram-based features are extracted from handwritten Devanagari word images, following the approach described in (Koerich, 2003). The contour of the handwritten word images is represented by the pixels of the outer and inner boundaries, as illustrated in Fig. 4.3. This representation can be obtained by examining each pixel within a 3×3 window frame.



(a) **(b)**

Figure 4.3: Examples of contour extracted from Devanagari handwritten word images: **(a)** "ऋषिकेश" (Rishikesh) and **(b)** "अहमदाबाद" (Ahmedabad)

Further, the resultant contour is split into 64 parts corresponding to the 64 zones (4 × 16) as given in the Fig. 4.4.



(a) **(b)**

Figure 4.4: Splitting of contour into 64 parts corresponding to the 64 zones: **(a)** "ऋषिकेश" (Rishikesh) and **(b)** "अहमदाबाद" (Ahmedabad)

To obtain the directional histogram, the contour is traversed for each of the mentioned zones using windowing techniques. Zoning is employed to capture local properties or features instead of global features, thereby enhancing the recognition accuracy. In this work, the combination of gradient and structural features is also investigated to study their impact on the recognition accuracy of the Handwritten Word Recognition (HWR) system. The feature vectors obtained from the gradient-based and structural-based feature extraction techniques mentioned earlier are concatenated to construct a new feature vector.

4.3 CLASSIFICATION TECHNIQUES

Classification techniques are used to recognize the class of unknown word based on the extracted feature vectors. To recognize the unknown word, initially classifier is trained using various training samples. Thereafter, features shall be extracted from testing samples and compared with that of the training sample features so as recognize the class of unknown word. In this work, three classifiers namely Support Vector Machine (SVM), Naive Bayes (NB) and eXtreme Gradient Boosting (XGBoost) have been explored for the recognition of handwritten Devanagari words. The performance of these classifiers for recognition of handwritten Devanagari words have been analyzed in terms of various performance metrics based on the feature vectors obtained from from the gradient-based and structural-based features extracted from images.

4.4 EXPERIMENTAL RESULTS AND DISCUSSION

To examine the performance of handwritten Devanagari word recognition system, experimentation has been carried out using Python platform installed with various machine learning libraries such as Keras and Tensorflow etc. Due to the lack of standard dataset (Singh et al., 2022a), in present work, a corpus of 20,000 collected handwritten Devanagari words have been considered as dataset. The dataset is divided into 14,000 training words (70% of dataset) and 6,000 testing words (30% of dataset). System performance in terms of different metrics such as Recognition Accuracy (RA), False Acceptance Rate (FAR), False Rejection Rate (FRR) and Precision (PR) are given in Tables 4.1 to 4.4. Graphical representation of system performance are also presented in the Figs. 4.5 to 4.8.

4.4.1 System Performance based on Recognition Accuracy (%)

Table 4.1 presents the system performance based on recognition accuracy (%) for different feature extraction techniques and classification techniques. For the Gradient feature extraction (FA1) technique, the highest recognition accuracy is achieved with the eXtreme Gradient Boosting (XGBoost) classifier at 87.51%, followed by Naive Bayes (NB) at 85.48%, and Support Vector Machine (SVM) at 82.40%. In the case of the structural feature extraction (FA2) technique, again, the highest recognition accuracy is obtained with XGBoost at 89.74%, followed by NB at 86.73%, and SVM at 83.10%. When combining both gradient and structural features (FA1+FA2), the recognition accuracy improves further. XGBoost achieves the highest accuracy at 90.10%, followed by NB at 87.88%, and SVM at 85.73%.

Table 4.1: System performance based on recognition accuracy (%)

Feature Extraction Techniques	Classification Techniques		
	Support Vector Machine (SVM)	Naive Bayes (NB)	eXtreme Gradient Boosting (XGBoost)
Gradient Feature Extraction (FA1)	82.40%	85.48%	87.51%
Structural Feature Extraction (FA2)	83.10%	86.73%	89.74%
Combination of Gradient and Structural Features (FA1+FA2)	85.73%	87.88%	90.10%

In Fig. 4.5, the recognition accuracies in terms of percentage are presented visually. It can be observed that XGBoost consistently outperforms the other classifiers across all feature extraction techniques. This indicates the effectiveness of XGBoost in capturing the patterns and characteristics of the features, resulting in improved recognition accuracy. Overall, the results suggest that combining gradient and structural features leads to better performance compared to using each feature extraction technique individually.

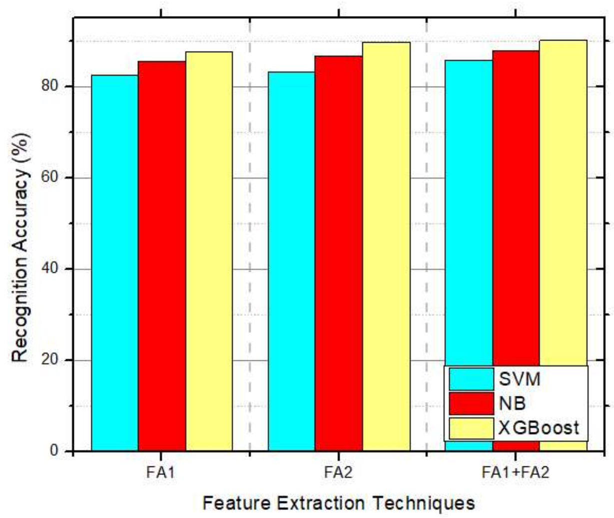


Figure 4.5: Graphical view of system performance in terms of recognition accuracy (%)

4.4.2 System Performance based on FAR (%)

Table 4.2 provides the system performance based on the False Acceptance Rate (FAR) for different feature extraction techniques and classification techniques. XGBoost consistently achieves the lowest FAR across all feature extraction techniques, with values ranging from 0.20% to 0.25%. Naive Bayes (NB) and Support Vector Machine (SVM) follow with slightly higher FAR values. The combination of gradient and structural features (FA1+FA2) leads to further improvement in FAR, with XGBoost achieving the lowest rate of 0.20%. These results highlight the effectiveness of XGBoost in minimizing false acceptances and its suitability for accurate predictions in the given recognition system. Graphical representation has been given in Fig. 4.6, which shows the performance comparison of different feature extraction techniques and classification techniques considered for this work, in terms of the False Acceptance Rate (FAR).

Table 4.2: System performance based on FAR (%)

Feature Extraction Techniques	**Classification Techniques**		
	Support Vector Machine (SVM)	**Naive Bayes (NB)**	**eXtreme Gradient Boosting (XGBoost)**
Gradient Feature Extraction (FA1)	0.35%	0.30%	0.25%
Structural Feature Extraction (FA2)	0.34%	0.27%	0.21%
Combination of Gradient and Structural Features (FA1+FA2)	0.29%	0.24%	0.20%

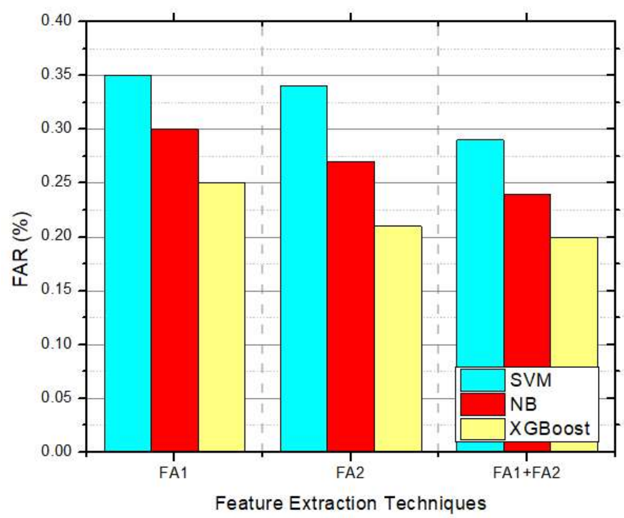


Figure 4.6: Graphical view of system performance in terms of FAR (%)

4.4.3 System Performance based on FRR (%)

In Table 4.3, the system performance is evaluated based on the False Rejection Rate (FRR) for different feature extraction techniques and classification techniques. The results show that XGBoost consistently outperforms the other classification techniques with the lowest FRR values ranging from 9.89% to 12.48%. Both Naive Bayes (NB) and Support Vector Machine (SVM) demonstrate higher FRR values across the feature extraction techniques. However, when combining the gradient and structural features

(FA1+FA2), the overall FRR is significantly reduced, with XGBoost achieving the lowest rate of 9.89%.

Table 4.3: System performance based on FRR (%)

Feature Extraction Techniques	Classification Techniques		
	Support Vector Machine (SVM)	Naive Bayes (NB)	eXtreme Gradient Boosting (XGBoost)
Gradient Feature Extraction (FA1)	17.60%	14.51%	12.48%
Structural Feature Extraction (FA2)	16.90%	13.26%	10.25%
Combination of Gradient and Structural Features (FA1+FA2)	14.26%	12.11%	9.89%

Graphical view of system performance as measured in terms of False Rejection Rate (FRR) is depicted in the Fig. 4.7. These findings emphasize the effectiveness of XGBoost in minimizing false rejections and highlight the potential of combining gradient and structural features for improved system performance.

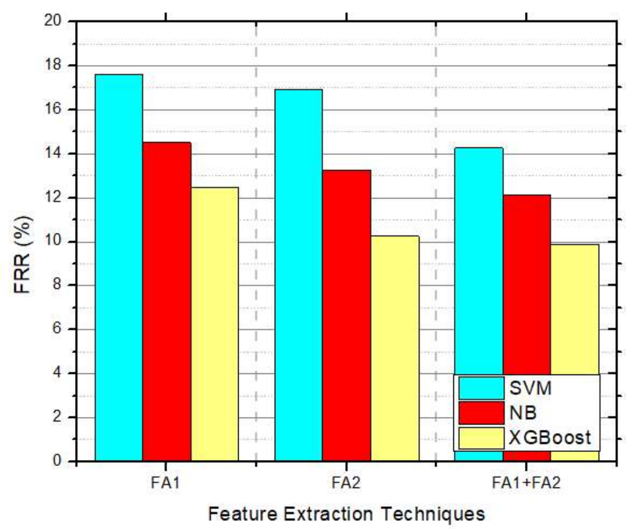


Figure 4.7: Graphical view of system performance in terms of FRR (%)

4.4.4 System Performance based on Precision (%)

Table 4.4 provides the analysis of system's Precision based on different feature extraction and classification techniques. The results demonstrate that XGBoost consistently achieves the highest precision values, ranging from 88.46% to 90.65%, across all feature extraction techniques. NB and SVM also yield competitive precision values, albeit slightly lower than XGBoost. The combination of gradient and structural features (FA1+FA2) further enhances precision, with XGBoost achieving the highest precision of 90.55%. These findings underscore the effectiveness of XGBoost and the benefits of incorporating both gradient and structural features for improved precision in Devanagari script recognition.

Table 4.4: System performance based on precision (%)

Feature Extraction Techniques	Classification Techniques		
	Support Vector Machine (SVM)	Naive Bayes (NB)	eXtreme Gradient Boosting (XGBoost)
Gradient Feature Extraction (FA1)	84.31%	86.90%	88.46%
Structural Feature Extraction (FA2)	84.78%	87.86%	90.65%
Combination of Gradient and Structural Features (FA1+FA2)	86.84%	89.04%	90.55%

Graphical view of system performance in terms of Precision (PR) is shown in Fig. 4.8. It show that using eXtreme Gradient Boosting (XGBoost) classification, maximum Precision (PR) of 90.65% has been obtained by structural feature extraction (FA2) technique.

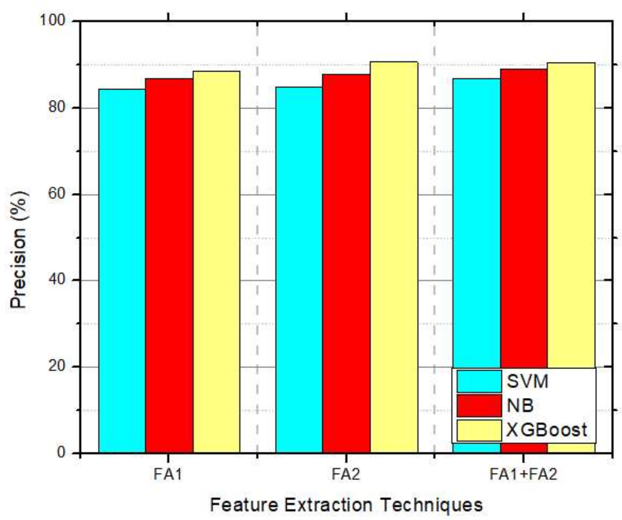


Figure 4.8: Graphical view of system performance in terms of PR (%)

4.5 ANALYSIS BASED ON EXPERIMENT RESULTS

The results indicate that combining both gradient and structural features leads to improved recognition accuracy compared to using them individually. Furthermore, the choice of classification technique also impacts the overall performance of the handwritten word recognition system. One of the confusion matrix has been depicted in the Fig. 4.9 using combination of gradient and structural features (FA1+FA2) and XGBoost classification. Class-wise performance for the proposed system of handwritten Devanagari word recognition can be analyzed using the confusion matrix.

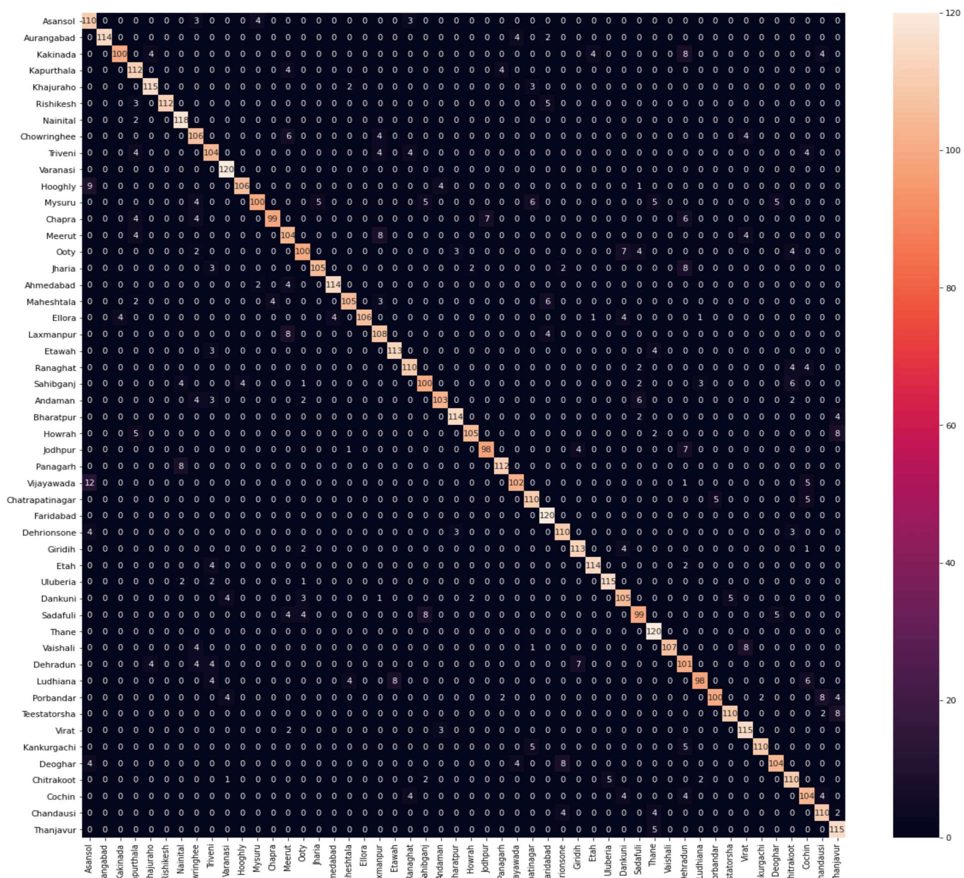


Figure 4.9: Confusion matrix plot

4.6 CHAPTER SUMMARY

This work investigates some techniques for holistically recognizing handwritten Devanagari words. Two features namely gradient as well as structural based features are studied using a corpus of handwritten words (Devanagari). Combination of these two features is also considered to improve the recognition accuracy of the proposed system. For recognition, three classification techniques namely Support Vector Machine (SVM), Naive Bayes (NB) and eXtreme Gradient Boosting (XGBoost) are used due to their robustness. A corpus of handwritten Devanagari words is collected from hundreds of writers belonging to various age groups, geographical backgrounds and qualifications. Using eXtreme Gradient Boosting (XGBoost) classification, a maximum Recognition Accuracy (RA) of 90.10%, minimum False Acceptance Rate (FAR) of 0.20% and False Rejection Rate (FRR) of 9.89% has been achieved with a concatenation of gradient and structural features. However, structural based features along with XGBoost classifier attained maximum Precision (PR) of 90.65%.

Overall, it has gathered from experimental work that combination of feature vectors resulted from gradient and structural features along XGBoost classifier perform better as compared with individual features itself. But, still there is a scope for improving the recognition results.