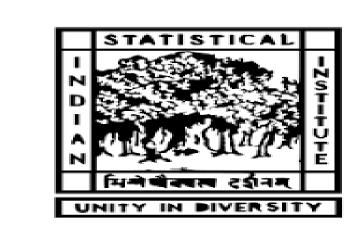


<sup>1</sup>Sauradip Nag, <sup>2</sup>Palaiahnakote Shivakumara, <sup>3</sup>Wu Yirui, <sup>4</sup>Umapada Pal, and <sup>5</sup>Tong Lu <sup>1</sup>Kalyani Government Engineering College, Kalyani, Kolkata, India

<sup>2</sup>Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia <sup>3</sup>College of Computer and Information, Hohai University, Nanjing, China

<sup>4</sup>Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India.

<sup>5</sup>National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China.

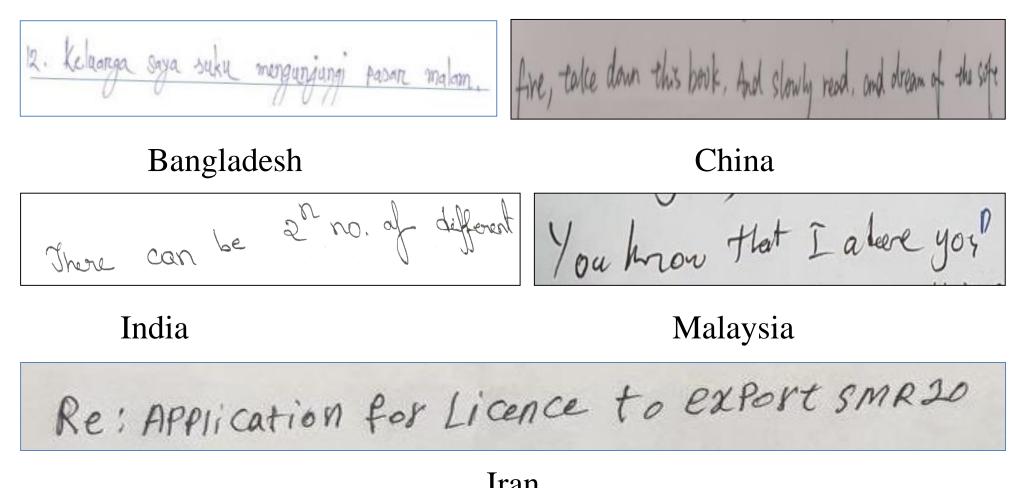


## Motivation

UNIVERSITY

OF MALAYA

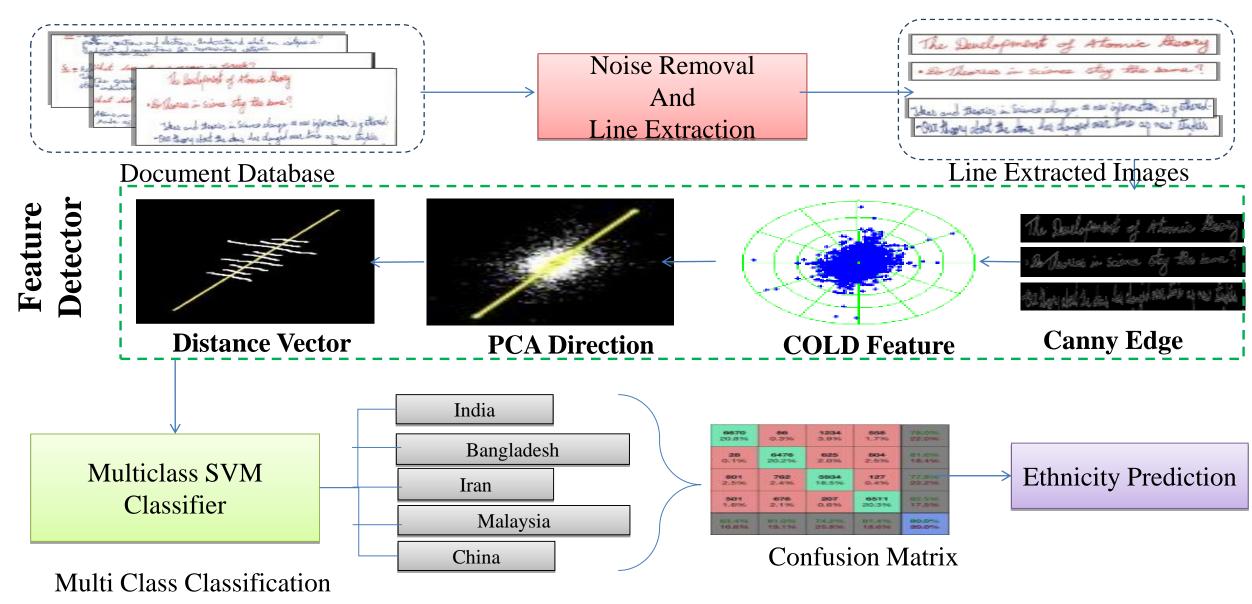
- Identifying ethnicity or nationality is useful for several crime applications where people of different countries are involved.
- There are methods for nationality and person identification using biometric features. However, these methods suffer from inherent limitations especially when the images are expose to open environment.
- Therefore, this work can help to assist forensic investigation to enhance the performance.



Examples of English handwriting of different countries

## Proposed Method

- This work propose to explores Cloud of Line Distribution (COLD) features for nationality identification using handwriting analysis.
- The proposed method uses tangent and mean intensity of edge pixels to segment the text line from the image and remove rule line.
- First, the proposed method finds dominant points on contour of the edge components using polygonal approximation.
- The distance between the dominant points are used for estimating polar coordinates, which gives COLD distribution in polar domain.
- The shape of the COLD distribution is used for nationality identification with the help of SVM classifier.



Framework of the proposed method

## Preprocessing for Text Components Detection

Horizontal Projection Profiles and Tangent angle for the edge pixels are computed as follows.

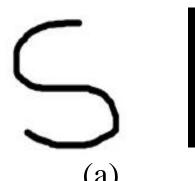
## **COLD** Distribution for Feature Extraction

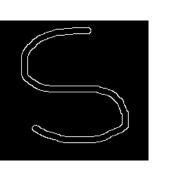
Preprocessing step for feature extraction.

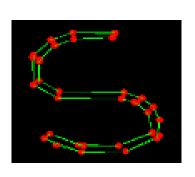
Polar coordinates are generated using distance between dominant points of the contours of edge components.

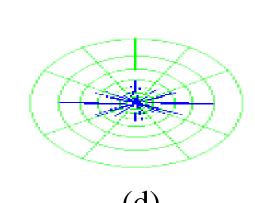
$$\theta = \tan^{-1} \left( \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right) \quad r = abs \left( \sqrt{(y_{i+1} - y_i)^2 + (x_{i+1} - x_i)^2} \right)$$

Here  $(x_i, y_i)$  and  $(x_{i+1}, y_{i+1})$  denote the coordinates of a dominant pair. A line segment can be represented using  $\theta$  and r as a point  $(\theta, r)$  in polar domain

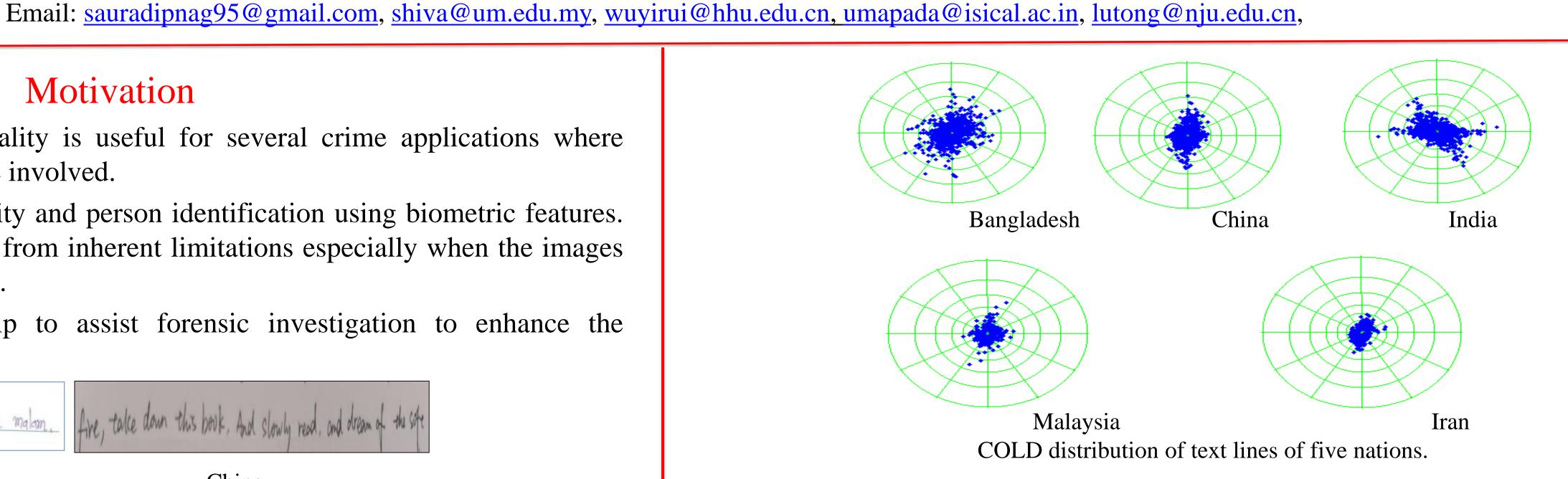






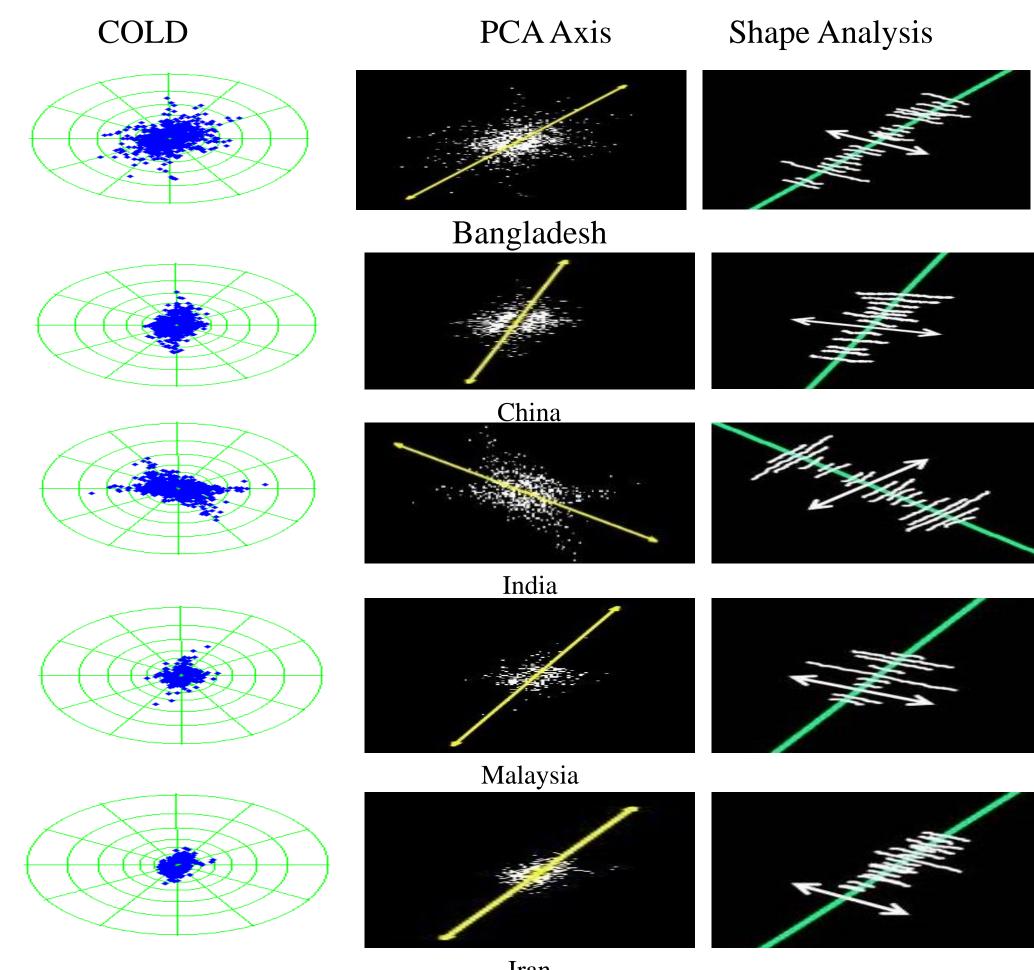


COLD distribution for handwriting components. (a) is a handwritten character, (b) is the Canny edge image of (a), (c) shows the dominant points for the contours, and (d) gives Cloud of Line Distribution (COLD) in polar coordinate.



#### Feature extraction for Nationality Identification

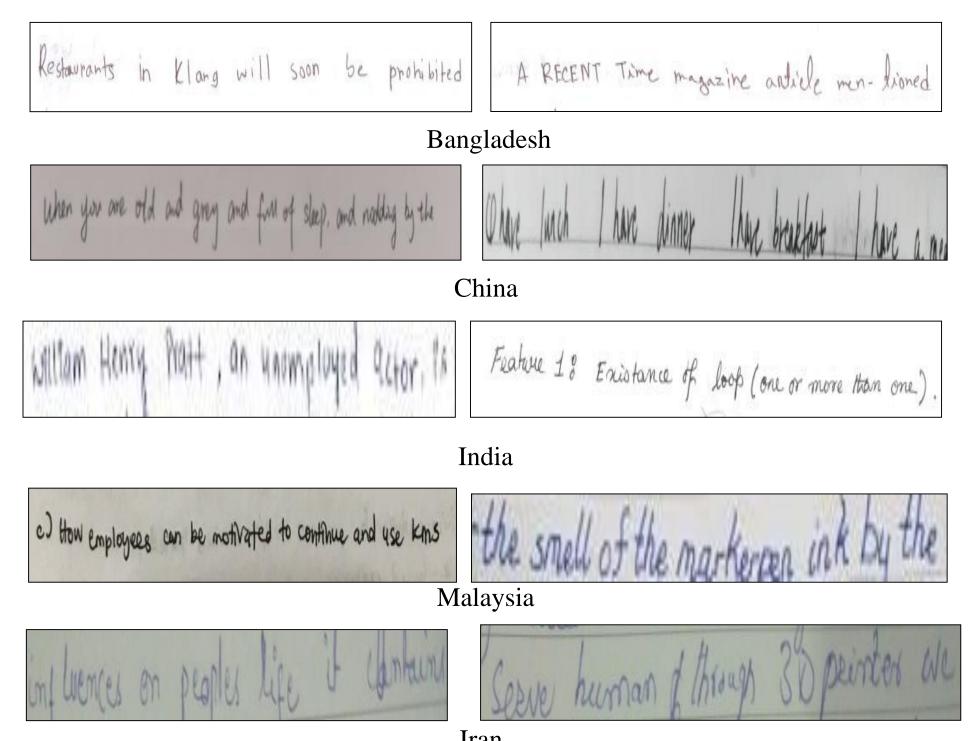
- The proposed method draw principal axis for the COLD distribution and calculate the distance from the pixels of principal axis to white pixels in the distribution.
- The absolute difference between reference pixel and pixel of distribution is computed and considered as features.



Feature extraction for handwriting text lines of five nations.

# **Experimental Results**

• Each class contains 100 images so total 500 images are used for experimentation, which includes 90% images written by male and 10% are written by female.



Sample of the successful handwritten text line images of the proposed method for five ASIAN countries.

## Confusion matrices of the proposed and exiting method (Maadeed et al.)

Classes	Bangladesh	India	China	Iran	Malaysia	Classes	Bangladesh	India	China	Iran	Malaysia
Bangladesh	79%	14%	0%	4%	3%	Bangladesh	41%	18%	10%	11%	20%
India	15%	73%	4%	6%	2%	India	21%	37%	15%	13%	14%
China	3%	7%	82%	3%	5%	China	12%	20%	38%	12%	18%
Iran	7%	13%	3%	66%	11%	Iran	19%	17%	9%	45%	10%
Malaysia	6%	10%	7%	2%	75%	Malaysia	17%	10%	23%	18%	32%
CR in (%)			75			CR in (%)			38.6		

# Conclusion

- We have proposed a new method based on COLD distribution in polar domain for ethnicity or nationality identification using handwriting analysis.
- The principal axis is used for feature extraction and SVM classifier is used for classification
- Next, our plan is to identify persons of different provinces within country where multilingual is official language.