# INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

**A PATH FOR HORIZING YOUR INNOVATIVE WORK**

## GENETIC ALGORITHM APPROACH FOR DEVNAGARI CHARACTER RECOGNITION

**S. P. RAMTEKE[1], P. M. WAGH[1], N. P. PATIL[2]**

1. Department of Electronics and telecommunication, SSBT's college of Engineering and Technology, Bambhori, Jalgaon, Maharashtra, India.
2. Department of Computer Engineering, SSBT's college of Engineering and Technology, Bambhori, Jalgaon, Maharashtra, India.

**Abstract:** Today's world is established a huge amount of research work in the field of character recognition system. In past few years, the vast amount of innovations in research work related to the character recognition of printed as well as handwritten documents in various types of scripts. The existence of different types of feature extraction and classification techniques improves the performance of optical character recognition system in various application areas. A comprehensive list of references is reported which elaborates various feature extraction and classification techniques. Genetic algorithm is an unbiased optimization algorithm which makes parameter selection in an optimized way so as to obtain the global optimum. This paper investigates how genetic algorithm is applied for optical character recognition system. Genetic algorithms are used in Optical character recognition for feature selection optimization. Devnagari script is very popular in India for Marathi, Hindi, Nepali, Sanskrit, Konkani etc. Genetic algorithm is class of evolutionary algorithm which reduces the classification error in optical character recognition system.

**Keywords:** Genetic algorithm, Fitness function, Feature extraction, Character recognition

*PAPER-QR CODE*

**Corresponding Author: MR. S. P. RAMTEKE**

**Access Online On:**

www.ijpret.com

**How to Cite This Article:**

353

## INTRODUCTION

A Recognition process is a very challenging task among the many intelligent activities of the human brain system. The recognition can be defined as analysis of a specific object. This object can be voice, handwriting, fingerprint, gender etc. In the field of pattern recognition these objects are called as "patterns". The pattern recognition can be defined as the area of science concerned with discriminating objects on the basis of information available about them. This information is called "Features". The optical character recognition (OCR) is a field of pattern recognition that provides a full alphanumeric recognition of printed or handwritten characters at electronic speed by simply scanning the form. OCR is used in variety of applications such as banking, healthcare, invoice imaging, optical music recognition, automatic number plate recognition, digital libraries etc.

Problem definition and proposed solution

OCR has various forms such as printed as well as handwritten. Printed character recognition is easier than handwritten character recognition. But the problems arise in printed character recognition when the fonts and sizes of a character vary as shown in fig.1. Another problem is some characters having similar features as shown in fig.2.

S s S ▯ S S S S

Fig.1 Representation of character "S" in different fonts

| j , i | V, W | S, 5 | I ,L, T | B,3, 8 | F, E |
|---|---|---|---|---|---|

Fig.2 Characters pairs of similar features

The handwritten character recognition is more difficult than printed version. Handwriting is differs from person to person and Even handwriting of same person may differs from time to time. As the handwritten characters may change with the mind state of the person. Thus, the powerful feature extraction and classification methods must be used for OCR system.

The devnagari script is used in various documentation. It includes Marathi, Hindi, Nepali, Sanskrit, Konkani etc. As Hindi is the most popular in India which uses devnagari script. Thus, this paper gives the methodology for devnagari script. These are some devnagari vowels.

**A Aa [ [- { {] e eo Aao AaOO AM A:**

Thus, the proposed work is designed an optical character recognition system using genetic algorithm for handwritten character recognition for devnagari script.
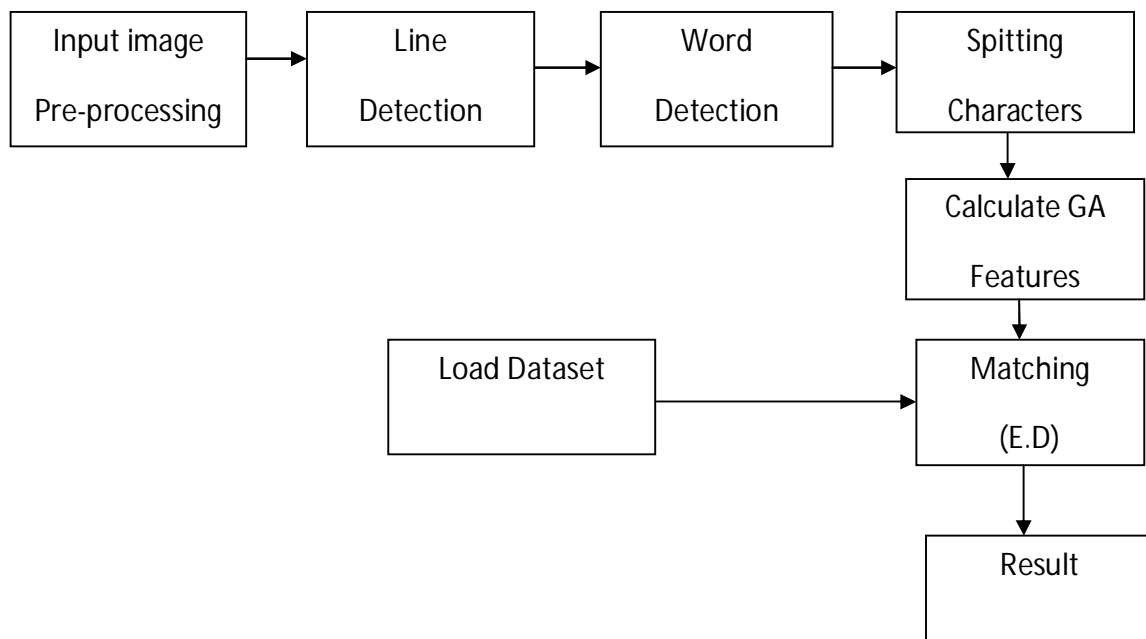
Fig.3. system model

III. PROPOSED SYSTEM MODEL

Fig. 3 shows the proposed system model for optical character recognition system it includes all the basic steps of every character recognition system i.e. pre-processing, segmentation, feature extraction, classification and post processing. Input image is pre-processed and given for segmentation. Segmentation is done by taking vertical and horizontal projections from histogram. It has 3 parts splitting line, word and characters. The next step is feature extraction which selects and prepares data which is used by a classifier to get the recognition task. The

355

genetic algorithm is used for this task. Next, matching in which the comparison is done between the input database and output of feature extraction stage. The performance of classifiers is dependent on features. The Euclidean distance is used for matching in dataset and the calculated features. Finally, the result is displayed which is a post processed image. The MATLAB framework is used for implementing the system.

IV. METHODOLOGY OF GENETIC ALGORITHM

Though various methods have been proposed for OCR system the use of GA is increased rapidly. Genetic algorithms offer a particularly attractive approach for this kind of problems since they are generally quite effective for rapid global search. A GAs is an optimization and search technique utilized in computer science to find approximate solutions to problems. It is inspired by processes in biological evolution such as natural selection, inheritance, recombination, and Mutation.
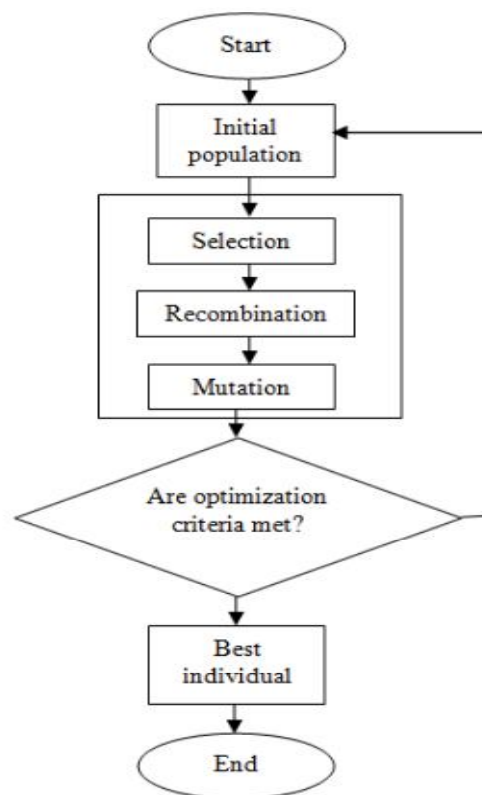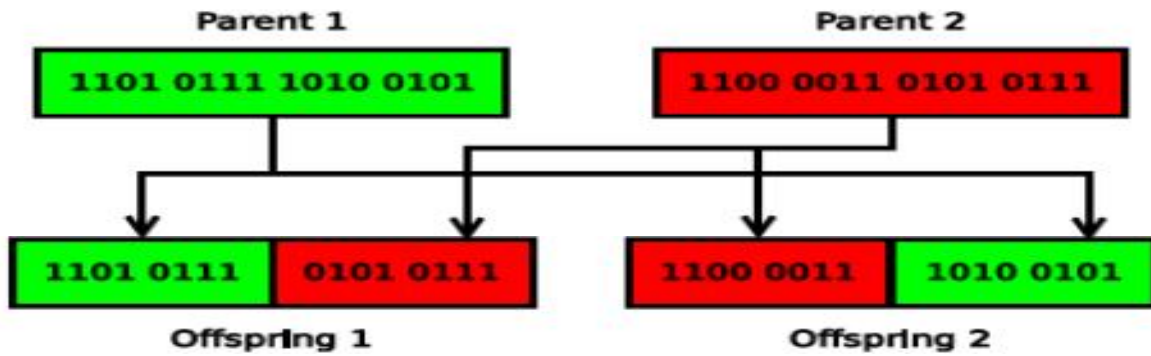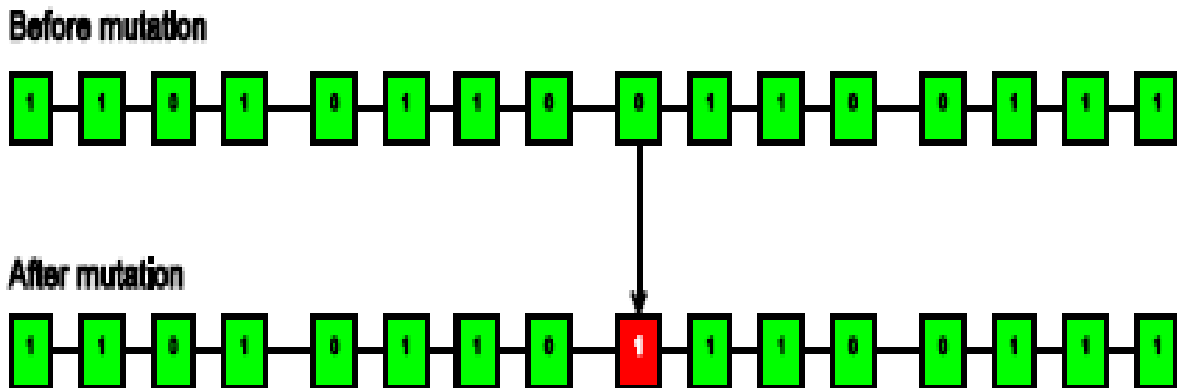


Fig.4 Flow chart of GA [1]

Fig.5 Crossover [9]



Fig.6 Mutation [9]

## A. *Initial Population*

Initially many individual solutions are (usually) randomly generated to form an initial population. The population size may differ due to the nature of problem. A population consists of 'N' individuals where every individual consists of a chromosome having 'L' genes. The following notation will be used to describe the chromosome of individual number i

Where 1<=i<=N;

Chromosome I= gene1....................L.

## B. *Evaluation*

The evaluation is done to calculate fitness value of each individual. The fitness function is defined over the genetic representation and measures the quality of the represented solution. The fitness function is dependent on problem domain. How to calculate the fitness value is an

357

important choice since that the fitness value decides which individuals will get to pass on its genes. The value returned from the fitness function for one gene represents the degree of matching between the feature vector of the character represented by that gene and the feature vector of the real character.

## C. *Selection*

In the selection, we add up the fitness of all the individuals and then calculate how much each individual contributed to the total fitness of the population. During each successive generation, a proportion of the existing population is selected to breed a new generation. Select specific individuals from the population to be the parents that will used to create new individuals.

## *Crossover*

The cross over is the process of creating offspring from the parents. It is also called as recombination of offspring from the parents. A single static point cross over is used to generate offspring from the selected parents we use. The offspring 1 is created from first part of parent 1 and later part of parent 2. Similarly, the offspring 2 is created from first part of parent 2 and later part of parent 1. Figure 5 illustrated this process of crossover. The crossover can be 1-point, 2-point. After crossover mutation process is carried out.

## D. *Mutation*

In mutation a bit in individual offspring is flipped from $0 \rightarrow 1$ or $1 \rightarrow 0$ at random. Mutation is done to avoid the recombination of the same feature vectors and to expand the explored solution space, as well as to alternate between some chromosomes.

Fig.6 shows the process of mutation which flips the 0 bit into 1.Every chromosome is simply scanned gene by gene and with a mutation rate Pm. The probability of mutation usually kept small i.e. Pm =1/L.

## CONCLUSIONS

In pattern recognition areas, the character recognition is most popular and challenging field. Due to various problems faced by OCR system an innovative and promising technique should be used. Thus, genetic algorithm is an optimized method for OCR and proves a better efficiency up to 90-95 %. The recognition rate is dependent on fitness function and population size.

## REFERENCES

1. Vedgupt Saraf, D. S. Rao," Devnagari Script Character Recognition Using Genetic Algorithm for Get Better Efficiency", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, April 2013.

2. K. R. Dahake, S. R. Suralkar, S. P. Ramteke," Optical Character Recognition for Marathi Text Newsprint", International Journal of Computer Applications Volume 62– No.16, January 2013

3. Nilima P. Patil, K. P. Adhiya, Surendra P. Ramteke, SSBT'S College of Engineering & Technology Bambhori, Jalgaon, "A Structured Analytical Approach to Handwritten Marathi vowels Recognition", International Journal of Computer Applications Volume 31– No.3, October 2011.

4. Raj Kumar Mohanta, Binapani Sethi," A Study on Application of Artificial Neural Network and Genetic Algorithm in Pattern Recognition", International Journal of Computer Science & Engineering Technology (IJCSET) Vol. 3 No. 2 February 2012.

5. R. Jayadevan, S. R. Kolhe, P. M. Patil, U. Pal," Offline recognition of Devnagari Script : A survey",IEEE transactions on systems, MAN, and Cybernetics- part C: Applications and reviews 2010.

6. S. Mori, K. Yamamoto, M. Yasuda, "Research on Machine Recognition Of Hand printed Characters", IEEE Trans. Pattern Analysis, Machine Intelligence, vol.6, no.4, pp.386-404, 1984.

7. M.Sriniwas,L.M.Patnik," Genetic algorithms: A survey", IEEE 1994.

8. Rahul Khadse," Survey of Research on Optical Character Recognition using Artificial Neural Network, Genetic Algorithm, Fuzzy Logic and Vedic Mathematics".

9. Emil Gedda & Kalle Lindqvist, "Genetically modelled Artificial Neural Networks for Optical Character Recognition  - An evaluation of chromosome encodings",  Master's Thesis Computer Science December 2011

10. M.A. Abed, Ahmad Nasser Ismail and Zubadi Matiz Hazi," Pattern recognition Using Genetic Algorithm", International Journal of Computer and Electrical Engineering, Vol. 2, No. 3, June, 2010 1793-8163.