# Optimization and Classification of Fruit using Machine Learning Algorithm

**Ms. Snehal Mahajan**
*Student*
*Vishwakarma Institute of Technology, Pune, India*

**Prof. S. T. Patil**
*Professor*
*Vishwakarma Institute of Technology, Pune, India*

## Abstract

In the world of Automation there is agriculture which is come into play to increase productivity, quality as well as economic growth of the country. Fruit classification is an important process for separating different fruits. For this purpose support vector machine (SVM) and Genetic algorithm (GA) is using to give best result. Here GA and SVM perform separate operation where GA is use for optimization of feature and then SVM is use for classification. Three type of fruit images i.e. apples, banana and grapes are using for classification. 178 fruit images is using for series of experiment. Feature extraction process performed after preprocessing process. In feature extraction process, shape, texture and color of image characteristics extracted. Than above optimization process is use for improving classification process and get better accuracy.

**Keywords: Fruit classification, Feature optimization, Feature extraction, Genetic Algorithm, Support vector machine**

## I. INTRODUCTION

Each and every data has its some special features. In all over the world there are different kind of datasets like music datasets, people, trees, fruits, etc. all kind of datasets has their own special features. Every fruits has their own special features like shape, texture, color which helps to find the every fruits identity. So to extract these features from given fruits datasets is most important part in fruit classification.

It is required to construct an automatic fruit classification system in various fields. It is necessary to select appropriate feature from extracted feature so that unwanted feature won't use and classification process perform on best selected features only. The use of all features without selecting the best features cause error in classification [1] and increase the time required to complete the process [2]. In real life application fruit recognization and classification is used many times. It can be used for children as an interactive tool to improve their learning talent and used in supermarket checkout system for manual barcodes [3], [4]. It can use in fruit machine to get the particular fruit according to user requirement [4]. People who has eye weakness problem used it as an tool for fruit identification [8].

The aim of this paper is to investigate the use of GA and SVM for optimization and classification. The proposed system includes feature extraction, feature optimization and classification. The paper structure is as follow. Two proposed technique of classification is introduced in section II. Feature extraction part is discourse in section III. The section IV describe about GA use for feature optimization. Fruit classification using SVM in separate method is described in section V. Section VI describes the combine method of GASVM for classification. VII section tells about the result of experiment. Finally, VIII section discourse about conclusion of the experiment.

### A. Problem:

1) *K-Nearest Neighbor:*

Process of this classification is based on nearest training sample of the features space data. The process of training for k-NN is to store some feature vector and label it as a training image. The remaining data is use for testing.

2) *Limitations of k-Nearest Neighbor:*

k-NN is lazy learner which means it use direct training data for classification and do not use it for learning purpose. It find the distance and sort the data which perform slow process.

3) *Neural Networks:*

In the format of NN each node is connected to each other which consist of weight. The priority of connection is designed according to the weight of connection.

### B. Limitations of Neural Networks

The train data in NN application cannot retrain i.e. if new data is wanted to added later than its equivalent to impossible to add new data in existing system. In NN time series handling data is very difficult.

## II. THE PROPOSED FRUIT CLASSIFICATION USING TWO DIFFERENT TECHNIQUE

- Feature extraction: Form fruit images three main types of features are extracted which are color, texture and shape.

– Optimization and classification stage: The extracted features are used in optimization. GA is used for optimization purpose and SVM is used for classification purpose.

## III. FEATURE EXTRACTION

Color features: here mean of all the RGB is calculated.

$$M = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{x_{ij}}{nm}$$

Where,: pixel value of ith row and jth column and Image is of size m x n, ij x

Shape features : Eight different features are extracted here, which are area , perimeter , roundness, equivdiameter, convex area, major length, minor length and eccentricity.

The total no. of pixels inside the object is considers as an area. The perimeter is calculated by using distance between each pixel in the boundary. Roundness and equivdiameter is calculated by using area and perimeter. Convex area, major length, minor length and eccentricity are calculated by using the image property.

Texture feature : standard deviation and variance of RGB values are calculated. Variance is average of squared difference from mean. It is calculated as follows:

$$\text{var} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - M)^2$$

Deviation means how far from normal. Standard deviation is a measure of how spread-out numbers are.

$$\sigma = \sqrt{\text{var}}$$

## IV. FEATURE OPTIMIZATION USING GENETIC ALGORITHM

All extracted features from an image used in this process where it selects the optimal features of the images. So the unwanted and unused features removed. Here selected features used as an input and then calculate its fitness function. The initial population is generated randomly by default and next generation of the population is generated by using the fitness of the individuals in the current generation. The process stops when the maximum number of generation is reached or when there is no change in the best fitness value from time given in seconds. GA is one of the best algorithms to work on large scale problems to solve it on optimal or near optimal solution. The GA works are sets as follows:

### *A. Initial Population:*

Initial population consists of all random feature vectors which are called as chromosomes. These chromosomes are most important part in genetic algorithm. Genetic algorithm has the ability to handle large number of population [7].

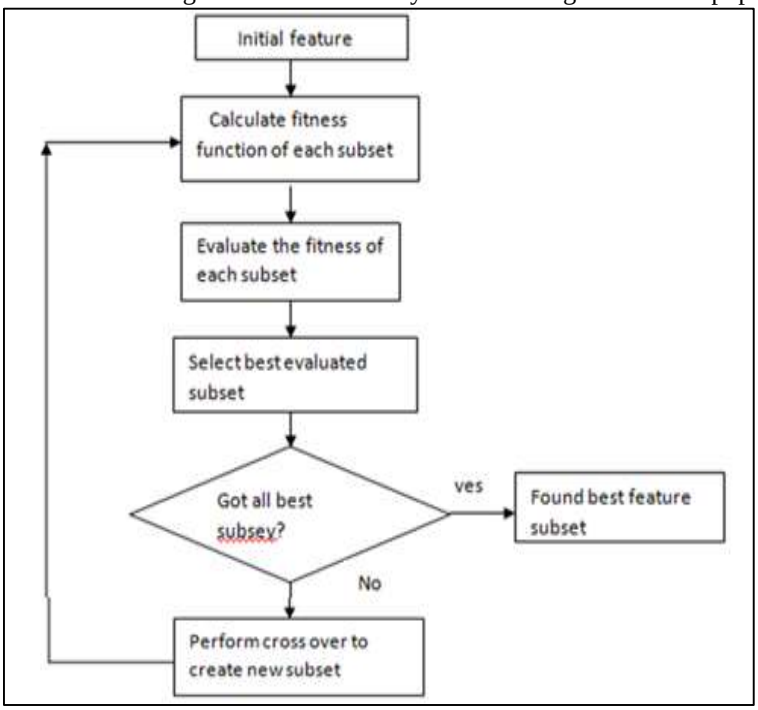


Fig. 1: Flow chart for optimal feature selection of GA

### *B. Fitness Function:*

Fitness function is most important task in Genetic algorithm to perform feature optimization task.. Each individual has their own fitness value which calculates by using curtain rules and then finds that how close it is to the actual solution. Simple fitness function method used in it[6].
To find the fitness value of the each chromosomes it required encode the each feature value where it convert feature value into binary form and then select 1's and 0's position not been selected.

$$f = 100 * ((x_1)^2 - x_2 ..... - x_n)^2 + (1 - x_1) + (1 - x_2) + ... + (1 - x_n)$$

### *C. Selection:*

In selection tournament function use to select best two features. It select the best two minimum value of fitness features. Here they apply complete learning rule and these two features elite and goes into next generation [7], [8].

### *D. Recombination:*

This process is different from reproduction. Its main aim is to save good solution and make its copy of it and then discard the bad solution in given population, so that the size of the population does not change. Here crossover and mutation processes perform for recombination [8].

## V. CLASSIFICATION USING SVM

In machine learning algorithm svm is supervised learning algorithm which is used as one of the best classifier and pattern recognizer in classification process. SVM is used for the classification of linear as well as nonlinear datasets. SVM provides sample of data inform of point which is place in space. Then separate in different classes in such a way that the width between two classes of margin is as possible [9][10].

$$w \bullet x + b = 0$$

The equations of the hyper-planes selected by the SVM are given as:

$$\vec{w}.\vec{x} + b = -1$$
$$\vec{w}.\vec{x} + b = +1$$

That mean in linear svm we required hyperplane where we find the maximum margin of hyperplane and that margin is built by samples of two classes. The sample of that margin is called support vector [10]. Once we train our data we use support vector to classify test touple.

$$f(\vec{x}) = sign\ (\sum_{i=1}^{N} \alpha_i y_i \vec{x}_i . x_i + b)$$

Initially svm was built to solve problem of two classes only [9], but it can extend for more number of classes. For that they introduce multiclass svm. In the process of multiclass SVM one vs All concept is used, where any one class is consider as +1 class and remaining are consider as -1. In this way all classes get trained. At the time of testing process it find the probability of that touple to place in which class.

For the nonlinear data kernel function is use. The process of non-linear classifier is usually similar to liner process, except that it replaces each and every dot into nonlinear kernel function. So that the svm algorithm can fit in to the maximum margin hyperplane and transform the features space. So the kernel function is used to evaluate data in complex division of space. Radial basis kernel function is used for the classification purpose.

The difference between linear SVM and RBF kernel is that RBF kernel converts two-dimensional plane into higher dimensional plane [13], [14]. The radial basis kernel function represented as:

$$K_{RBF}(x, x') = \exp[ - \| x - x' \|^2 / 2\sigma^2 ]$$

Here $\sigma$ is std deviation and $\sigma^2$ is commonly called as variance. From variance we get the kernel parameter which is represented as $\gamma = 1 / 2\sigma^2$. So the equation will become:

$$K_{RBF}(x, x') = \exp[ -\gamma \| x - x' \|^2 ]$$

After the process of radial basis function kernel further all classification process is same as linear SVM classifier.

## VI. EXPERIMENTAL RESULT

Matlab R2012a is used for system implementation. The two different types of fruits datasets is used. The 70% of data is used for training purpose and 30% of data was used for testing purpose. All the original images have used after resizing for feature extraction purpose. The selected image of fruit is used to identify the similarity and differences between the color, texture and shape of the fruits. The min and max values of features for particular fruit is:

Table – 1
Feature extraction by Shape

| Fruit | Area | | Equivdiameter | | Perimeter | | Roundness | |
|---|---|---|---|---|---|---|---|---|
| | min | max | Min | max | min | max | min | max |
| Apple | 12407 | 29156 | 24.0807 | 192.6722 | 694.6 | 826.1 | 0.0361 | 0.3626 |
| Banana | 939 | 7450 | 17.9125 | 94.3942 | 328.707 | 528.753 | 0.0039 | 0.1121 |
| Grapes | 4106 | 24526 | 72.3044 | 176.7130 | 551.247 | 136.568 | 0.1933 | 0.5368 |

GA used to get optimal values of features. Total 178 fruits images are there to evaluate. The process of color, texture and shape feature extraction total 11 features extracted from each image. So total 1958 number of features are there. Sometime maximum number of features reduces the classification accuracy. Different combinatorial set of features should be obtained in order to keep the best combination to achieve optimal accuracy. As such, a GA-based feature selection will be used to reduce the number of features.

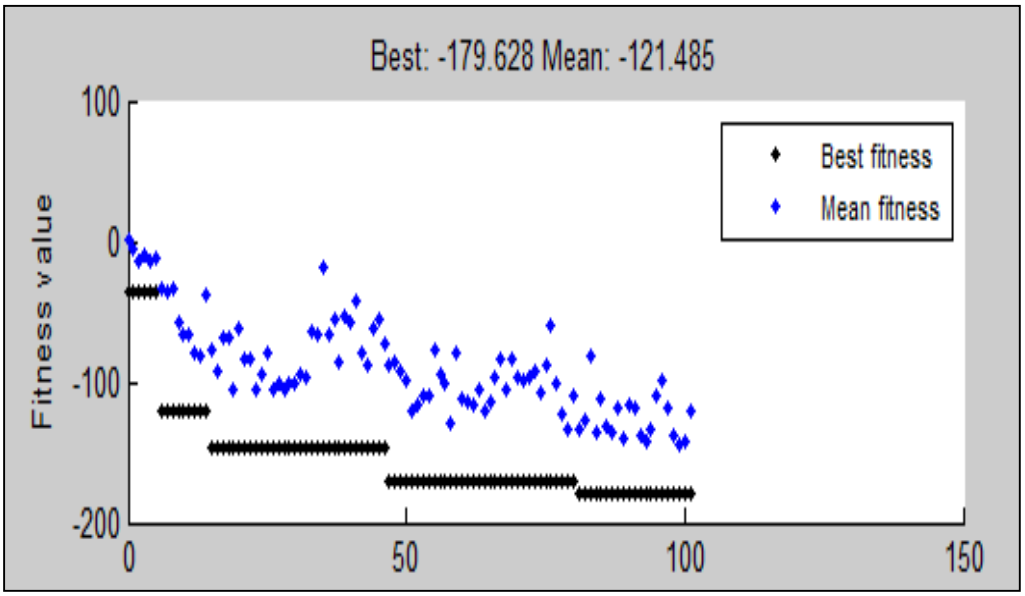


Fig. 2: GA simulation diagram

Once get the optimal features than perform classification process using SVM on all type of selected fruits. After completion of optimization process SVM process gives 96.77% accuracy result. Apple got 96-98Z accuracy and grapes got 95-97% accuracy in GA & SVM process. Banana got 96-97% accuracy.

## VII. CONCLUSION

This paper has presented optimization and classification using two algorithms GA and SVM. The proposed system contains feature extraction, optimization and classification. For feature extraction stage texture, shape and color of fruit feature extracted. The optimization process use to select the best features from total number of features. This process overcomes the disadvantage of SVM to take maximum time for training the data. So the optimization process helps to reduce time to complete classification process.GA select better features which help to classify fruits using SVM. This process gives 96.77% efficiency.

## REFERENCE

[1] Zawbaa H. M. , Hazman M., Abbass M. and Hassanien A. E. , "Automatic fruit classification using random forest algorithm", Hybrid Intelligent Systems (HIS) 4th international conference, Kuwait, 14-16 Dec. 2014, Pp: 164 - 168 .

[2] Haidar A., Haiwei Dong and Mavridis N. "Image-Based Date Fruit Classification" Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT), St. Petersburg, 3-5 Oct. 2012, pp 357–363.

[3] A. Rocha, D. C. Hauagge, J. Wainer and S. Goldenstein (2010, jan) "Automatic fruit and vegetable classification from image" vol 70 (Issue 1), pp. 96-104, Available : http://www.sciencedirect.com/science/article/pii/S016816990900180X.

[4] W.C. Seng and S.H. Mirisaee "A new method for fruits recognition system,"In Electrical Engineering and Informatics, ICEEI'09 International Conference, Selangor, 5-7 Aug. 2009, pp. 130-134.
[5] Rodrguez-Pulido F. J. , Gordillo B., Gonzlez-Miret M. L. and Heredia, F. J. (2013, nov) "analysis of food appearance properties by computer vision applying ellipsoids to colour data". Computers and Electronics in Agriculture. Vol- 99(issue 1), pp: 102-115, available : http://www.sciencedirect.com/science/article/pii/S0168169913002056
[6] Hirano Y. and Nagao T. "Feature transformation using filter array for automatic construction of image classification", Computational Intelligence and Applications (IWCIA), Hiroshima, 7-8 Nov. 2014, Pp: 59 – 64.
[7] H. Shahamat; A. A. Pouyan (2015, jan) "Feature selection using genetic algorithm for classification of schizophrenia using fMRI data", Journal of AI and Data Mining, vol. 3(issue 1), pp 30-37, Available : http://jad.shahroodut.ac.ir/article_331_70.html.
[8] Feng Tan, Xuezheng Fu, Yanqing Zhang, Anu G and Bourgeois (2008, jan) "A genetic algorithm-based method for feature subset selection", vol 12 (issue 2) Soft Comput, vol 12, pp 111–120, Available : http://link.springer.com/article/10.1007%2Fs00500-007-0193-8.
[9] Seema, A. Kumar and G. S. Gill "Automatic Fruit Grading and Classification System Using Computer Vision: A Review" Advances in Computing and Communication Engineering (ICACCE), 2015 Second International Conference, Dehradun, 1-2 May 2015 pp 598 – 603.
[10] Suresha M, Shilpa N. A and Soumya B (2012, aug) "Apples Grading based on SVM Classifier" International Journal of Computer Applications (0975 – 8878) on National Conference on Advanced Computing and Communications – NCACC, pp 27-30, Available : http://www.ijcaonline.org/proceedings/ncacc/number1/7993-1009
[11] Rocha, D. C. Hauagge, J. Wainer, and S. Goldenstein (2010, jan) "Automaticfruit and vegetable classification from images", Computers and Electronics in Agriculture, vol – 70 (issue 1), pp 96-104, availavle : http://www.sciencedirect.com/science/article/pii/S016816990900180X
[12] Ching-Tsorng Tsai, Hsian Min Chen, Jyh-Wen Chai, Clayton Chi-Chang Chen and Chein-I Chang "Classification of Magnetic Resonance brain images by using weighted radial basis function kernels" Electrical and Control Engineering (ICECE), 2011 International Conference on, Yichang, 16-18 Sept. 2011, pp 5784 – 5787.
[13] Fei Ma, Limin Yu, Mariusz Bajger and Murk J. Bottema "Mammogram Mass Classification with Temporal Features and MultipleKernel Learning" Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference, Adelaide, SA, 23-25 Nov. 2015 , pp 1-7.