



PERFORMANCE ANALYSIS OF FRUIT CROP FOR MULTICLASS SVM CLASSIFICATION

Shameem Fatima¹, M. Seshashayee²

¹Research Scholar, Department of CS, GITAM (Deemed to be University),
Visakhapatnam, India.

²Assistant Professor, Department of CS, GITAM (Deemed to be University),
Visakhapatnam, India.

¹ksufatima@gmail.com, ²mseshashayee@gmail.com

Corresponding Author: **Shameem Fatima**

<https://doi.org/10.26782/jmcms.2020.08.00015>

(Received: May 22, 2020; Accepted: July 22, 2020)

Abstract

The research study aim to improve the performance of fruit quality by two approaches, first by applying kernel technique combined with specific classification method support vector machine (SVM) with error-correcting output codes for fruit categorization and then by cross validation . It is measured by analyzing the different mention kernel selection on color and shape features. Two coding design method such as one-vs.-one and one- vs.- all are examined with three commonly used kernel function linear, polynomial (cubic) and Radial Basis Function (Gaussian). The Experiment was conducted on fruit dataset created from fruit 360 dataset with six categories such as Apples, Avacados, Bananas, Cherrys, Grapes and lemons. The accuracy obtained for the fruit category with 98% accuracy was enhanced by the proposed method by the use of kernel technique selection resulted to 99%. However kernel choice highly depends on the parameter used for fruit categorization is introduced and discussed. The Experiments was carried out to find the best SVM kernel among linear, cubic and Gaussian for fruit categorization. The Experiment also focuses on evaluation process using cross validation methods kfold and hold out which resulted in a better accuracy for the classification model. The results show that the proposed method provides very stable and successful fruit classification performance over six categories of fruits. The coding design one- vs. - one performed better when compared to one- vs. - all with respect to accuracy and training speed.

Keywords : Multiclass SVM, ECOC, kernel technique, KFold validation.

I. Introduction

In the field of the agriculture industry, Automation results in increasing the quality of the product, along with the economic growth and production of the country. The export of fruit in the market and quality evaluation affected by sorting of the fruits. In the post harvesting process even though the sorting and grading is

Copyright reserved © J. Mech. Cont.& Math. Sci.
Shameem Fatima et al

performed by humans, it is time consuming, expensive hence automation of the fruits grading system is needed which in turn depend on fruit quality based on shape, size and color [I].

SVMs were extended to multiclass classification problem from binary classification and it is an ongoing research these days. Multiclass classification in recent times acquired ample attention in research community of computer vision and machine learning. Several recent works for dividing the multiclass problem into a set of binary class problems is studied in literature appropriate for remote sensing image classification which “enhances the fault tolerance when solving multi-class classification problem” [XXIII], classifying the stress Assessment level [VII].

The Current approach for multiclass SVM are of two types. The first is by construction and combination of several binary classifiers including one-vs.-one and one-vs.-all and the second directly taking data in one optimization formulation [VII].

The Goal of our study is examine how well fruit categorization can be classified using Support vector Machine with Error correcting output code (SVM-ECOC) Framework with various kernel techniques. We would like to investigate whether or not the use SVM-ECOC with kernel technique can improve the accuracy compared to the earlier study [XV] on the same Fruit360 dataset [VIII].

The contribution of the current research with respect to the previous work can be summarized as

- We investigate the kernel technique with coding design approach to improve the accuracy of fruit categorization.
- We investigate the evaluation approach on KFold Cross Validation and Holdout to improve the accuracy of fruit categorization.

The image classification is performed using SVM with ECOC framework with various kernel techniques and cross validation approach. The two coding design (One-vs.-One and one-vs.-all) is experimented for image categories. The experiment evaluates the best kernel technique and classifier performance for automation of fruit categorization.

Table 1: Multiclass SVM with different Kernel Technique & Coding Design

Sno	Purpose	Multiclass SVM with Coding method	Features	Dataset used	Kernel used	Findings
1	Fruit categorization (2020) [XV]	Multiclass SVM-ECOC(6 categories of fruits) with OVO	Color , shape	Fruit 360	Linear kernel	Classification accuracy 98%
2	Date fruit (mohammed 2015) [IX]	Multiclass SVM (4 type of dates) with	Color, Texture, Size &	800 (4 types *200 images)	RBF kernel 10 folds cross-	Classification accuracy more than

		OVA	Shape		validation (C- set automatically using Lib SVM)	98%
3	Fruit and branch identification (2014) [XI]	Multiclass SVM (fruit, branch, leaf and background) with OVO approach	R,G,B	87 images	RBF kernel ($c=100$, $\gamma = 0.250$)	Citrus Fruit recognition rate 92.4%,
4	Strawberry (2013) [III]	Multiclass SVM (normal, powdery mildew, shrinkage of fruit disease & uneven ripening disease strawberries) with OVA strategy	Color texture & shape		Radial basis kernel function.	-
5	Fruit & Veg classification (dubey and jalal (2012) [XVI]	Multiclass SVM with OVA coding design	Texture	Supermarket dataset	-	Classification accuracy of the up to 99%,
6	Fruit classification (Zhang and Wu(2012)[XXI]	Multiclass KSVM(with OVO coding design	Color, texture & shape		Gaussian Radial Basic kernel	Classification accuracy of 88.2%

II. Methodology

Image Database

The Database contains 5817 images includes 6 categories (Apples, Avocados, Bananas, Cherrys, Grapes and Lemons) taken from the Fruits360 Dataset [VIII]. The Table: 2 show the total number of image in each category with sub categories. Out of 81 distinct folder of fruit category, 12 folders used to design six fruit categories.

Table 2: No of Image for each category

Category	Apples	Avocados	Bananas	Cherrys	Grapes	Lemons
Sub Category	Apple Red & Apple Golden	Avocado unripe & Avocados Ripe	Bananas green & Bananas red	Cherry Red & Cherry Yellow	Grape Pink & Grape White	Lemon & Lime
Total	971	918	890	984	892	892

Feature Extraction using Visual Features (color and shape)

The Feature extraction is performed using visual features [XIV]. As fruits come in various color and shapes. The feature extraction is performed based on visual feature (external feature) such as color and shape. The fruit color and shape is extensively used measured visual descriptor. The color analysis of the fruit has been proven best technique to make decision to determine the quality of the fruit [XVII]. Image color is generally described in three color spaces consisting of three dimensions such as LAB, RGB and HSV. The proposed work uses LAB (where L refers Luminosity, A signifying the color between red-green axis and B signifies blue-yellow axis). The given image is converted from RGB into LAB, the statistical value (mean) is evaluated and combined to form a color feature. The shape feature obtained by conversion to gray scale followed by evaluating value of threshold and conversion into binary image. It is then evaluated for edges using sobel and bounding box [XV].

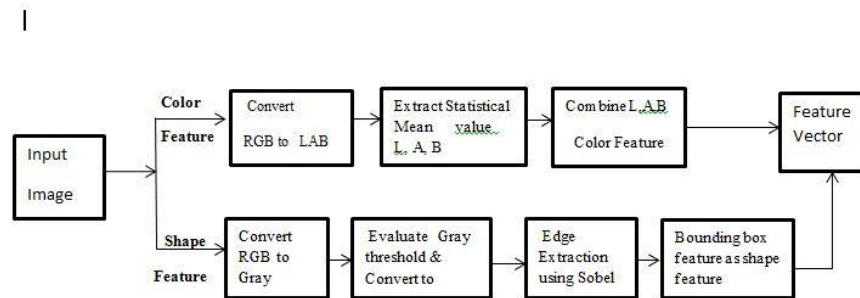


Fig. 1: Feature Extraction Process

Multiclass Support vector machine

In Supervised Learning, SVM is considered as one of the efficient classifier that has been widely used in image classification. Multiclass SVM is used for more than two category of classes. Dietterich et al (1995) proposed a framework in which ECOC (the error correcting output code) best suited to convert multiclass into many two class problems [XX]. The combination (ECOC with SVM) “enhances the fault tolerance of classification model when solving multiclass classification problem” [XXIV]. Svm is used for linearly separating hyper plane and it is extended to non-linearly separated hyper plane by the use of a technique called kernel function. It maps the data points into high dimensional feature space. Implementation of optimal kernel function is required for the classification problem in order to obtain optimal performance [XII].

For a nonlinear mapping ϕ that embeds input vectors into feature space, the kernels have the form:

$$K(x, y) = \langle \phi(x), \phi(y) \rangle$$

By choosing the K (kernel function) as a linear inner product, a polynomial or a radial basis function, the classification model becomes different as the SVM algorithms separate the training data in feature space by a hyperplane defined by the type of kernel function used [II].

The nonlinear function is of the form

$$f(x) = \text{sgn}\left(\sum_{i=1}^l (\alpha_i y_i k(x_i x) + b)\right)$$

The kernel functions that have been widely used are [XXIV].

Table 3: Types of Kernel Technique

Linear Kernel	$K(x, y) = (x \cdot y)$
Polynomial kernel with degree d	$K(x, y) = (1 + x \cdot y)^d$
Gaussian RBF kernel	$K(x, y) = \exp(-\ x - y\ ^2 / 2\sigma^2)$

III. Multiclass Fruit Categorization Algorithm

Input: Fruit Dataset

Output: Fruit category Accuracy Prediction

// Training

Input-Fruit Category Training Samples

Output-Acquire knowledge

Begin

For all Training Samples

Pre-Process the data

Train SVM with Kernel Technique w.r.t fruit categories

Evaluate Classifier with KFold Classifier

End for

End

//Prediction

Input- Fruit Image

Output- Accuracy Prediction

Begin

*Copyright reserved © J. Mech. Cont.& Math. Sci.
Shameem Fatima et al*

For the Fruit image

Extract color and Shape Feature

Apply SVM with ECOC Framework

Predict the classification result for the test image of Fruit Category

End for

End

IV. Performance Measures

The important aspect of an application is performance measure. There exists various performance metrics in literature for image classification depending on the technique used in various domains, different performance metrics are used for each domain. Accuracy is the measure that represents the category prediction. The confusion matrix shows the accuracy [V].

V. Experimental Results and Analysis

In our previous study [XV], we used two features to classify the fruits into categories to predict the quality of a fruit by its accuracy. This study investigates the selection of the kernel technique and coding design method which improve the accuracy. We proposed to choose best kernel technique that improves accuracy for the purpose, and its corresponding coding design method that reduces the training time. Using multiclass SVM with ECOC (error-correcting output code) along a suitable kernel technique, we successfully classified the fruits into categories with enhanced quality prediction. Improved accuracy was obtained with the selection of Gaussian kernel. The result shows that the training time is reduced by selection of the Polynomial kernel technique which is best suited for fruit categorization.

Evaluation of Model

Confusion matrix is used to evaluate the loss/highest accuracy. It shows the correct vs. incorrect labeled classes. The result shows that it is appropriate for image classification. The first twelve rows in Confusion Matrix signify the predicted class (Output Class) and the first twelve columns represent the true class (Target Class). The diagonal cells represent correctly classified categories. The off-diagonal cells represent incorrectly classified categories. All the categories and the category percentage are displayed in each cell.

Table 4: Confusion Matrix - Kernel Technique with one-vs-one

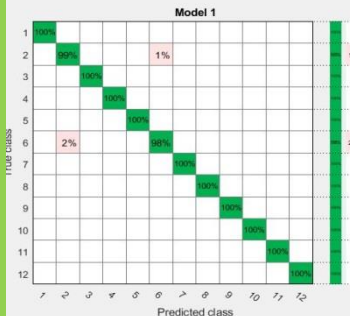
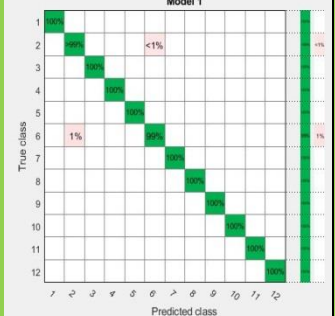
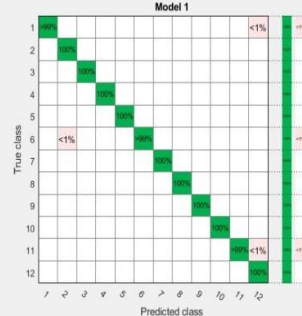
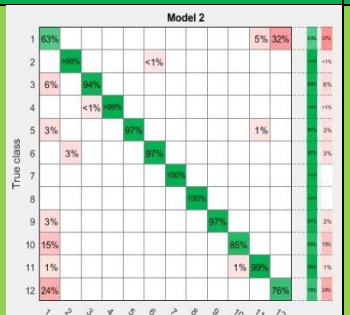
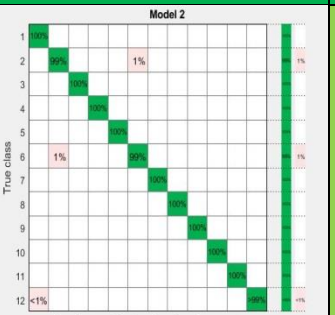
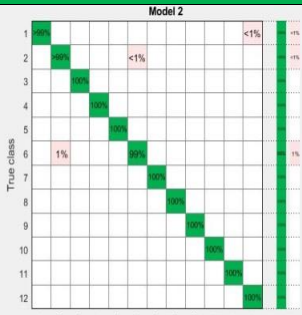
Linear Kernel	Polynomial kernel(Cubic)	Gaussian Kernel(RBF)
 <p>Accuracy 98-100%</p>	 <p>Accuracy 99-100%</p>	 <p>Accuracy >99-100%</p>

Table 5: Confusion Matrix - Kernel Technique with one-vs-all

Linear Kernel	Polynomial kernel(Cubic)	Gaussian Kernel(RBF)
 <p>Accuracy 63-100%</p>	 <p>Accuracy 99-100%</p>	 <p>Accuracy 99-100%</p>

The Table 4 & 5 presents confusion Matrix for kernel technique with one- vs.-one and one-vs. - all of each experiment. The result showed an accuracy of 98% with linear kernel, accuracy of 99% when polynomial kernel is chosen and the highest accuracy is achieved with Gaussian kernel. The accuracy shown by polynomial kernel is best suited for fruit categorization as it requires less training time.

Cross Validation Approach

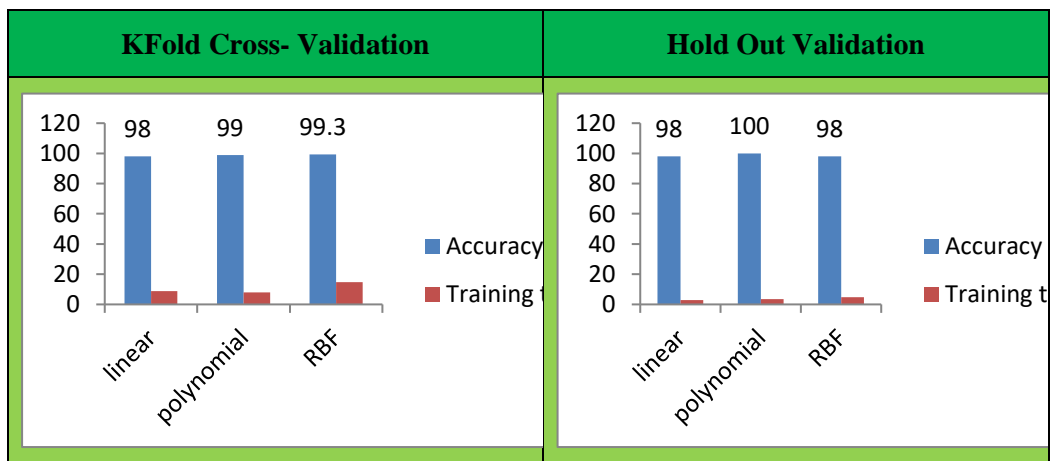
Cross-Validation is a considered as one of the statistical methods of evaluation by separating data into two sections, first segment is used for training a model and the other segment is used for validation. It typically follows the training sets and validation sets that treat cross-over in succeeding rounds such that each data point has a chance of being validated against [XIII]. There are two approaches that are most common and popularly used when a classification model is built and to test for accuracy are the KFold cross-validation and Holdout. Each approach is best applied in particular scenario.

KFold Cross-Validation and Hold Out

There are different validation methods available in the literature for sample selection as training data set. The KFold cross- validation method encompass the process of subdividing the actual samples into k equal sized sub-samples, each of it is considered as the validation data for testing of the classification model and the process is repeated k times. The value of k is not a fixed parameter which will be chosen by the user. One of the advantages of this method is over repeated random sub-sampling as training and validation for each for validation at least once [XVIII] [VI].

For evaluating efficiency of SVMs with selection of hyper-parameters, KFold cross-validation is commonly used. SVM with kfold cross-validation is treated as expensive because it requires k SVMs to train [XXII]. Holdout is an out of sample approach of evaluation wherein dataset is separated into a training-sets and a test-set. The test-set is thus out of sample data and sometime called the holdout data [IV].

Table 6: Experimental results – Cross Validation



VII. Discussion

The image categorization as multiclass problem has been applied with various frameworks with SVM in literature [X]. The research work chooses multiclass SVM with a framework of ECOC. It investigates the best method to improve the accuracy of the application with mention kernel technique using two commonly used coding design method (one vs one and one vs all) for training a classifier. For the K class problem results in K binary classifier. A construction of a set of K binary classifier each is performed to train a separate class from the rest and it combines them by performing the multiclass classification.

A similar research conducted by [XIX] based on the rice grain classification using image processing technique using multiclass SVM with two feature color and shape showed lower accuracy in comparison to the present research using multiclass SVM with one- vs.- one coding design approach.

The performance was improved by utilizing kernel technique with appropriate coding design. The experiment was conducted using MATLAB for six categories (with sub category), classification task produced improved performance with domain specific feature extractor and multiclass SVM with kernel technique.

Table 7: A comparison of result Accuracy

S.no	Ref	Classifier/Model	Kernel technique	Dataset	Features	Accuracy
1	[XIX]	Multi class SVM	-	-	Color and Shape	92.22%
2	[XV]	Multiclass SVM-	Linear	Fruit 360	Color ,Shape	98-100%
3	Proposed work	Multiclass SVM	Polynomial	Fruit 360	Color ,Shape	99-100%

VII. Conclusion

This paper explores the use of machine learning algorithm in the field of agriculture industry to evaluate performance of the multiclass classification model for fruit crop categorization. Among the most important quality characteristic feature of agricultural products such as color, shape, size and texture we have considered only the first two features for our research study. To replace manual inspection of fruit crop for the purpose of reducing post harvesting losses.

The evaluation of supervised classification is performed using a multiclass classification model based on kernel technique and cross validation approach is studied to improve the accuracy of the fruit categorization. The three kernel technique Linear, Polynomial and Radial basic Function is chosen for the evaluation purpose with two coding design methods such as one- vs. - all and one- vs.-one. The results indicate that proposed system of multiclass SVM with polynomial kernel technique with the one-vs. -one coding design performed best with combination of good

accuracy and desirable training time for quality prediction. The prediction accuracy of each category by confusion matrix is computed and presented.

This research also reports a comparative study on machine learning algorithms on cross validation approach to improve the accuracy for the classification model. Findings from this study showed that proposed multiclass SVM achieve better accuracy for KFold cross validation with 5 fold strategy and HoldOut need to further experimented as it is mainly used for large dataset. The experiment was conducted on fruit360 dataset which is the high quality fruit image dataset. The research can be used by farmers during sorting and grading process and the food factories can also benefit with fruit quality categorization techniques for the purpose of high quality maintenance, automatic packaging and to reduce labor cost. As a future work, we can add new features, investigate real time response method, analyze and optimization algorithm.

References

- I A. Bhargava, A. Bansal, Fruits and vegetables quality evaluation using computer vision: A review. Journal of King Saud University – Computer and Information Sciences (2018), <https://doi.org/10.1016/j.jksuci.2018.06.002>
- II C. Maaoui, & A. Pruski, (2008, July). A comparative study of SVM kernel applied to emotion recognition from physiological signals. In 2008 5th International Multi-Conference on Systems, Signals and Devices (pp. 1-6). IEEE.
- III C. Ouyang, D. Li, J. Wang, S. Wang, & Y. Han, (2012, October). The research of the strawberry disease identification based on image processing and pattern recognition. In International Conference on Computer and Computing Technologies in Agriculture (pp. 69-77). Springer, Berlin, Heidelberg.
- IV C. Sammut and G. I. Webb, Eds., “Holdout Evaluation,” in Encyclopedia of Machine Learning and Data Mining, Boston, MA: Springer US, 2017, p. 624.

- V Donahue, Jeff, et al. "Decaf: A deep convolutional activation feature for generic visual recognition." arXiv preprint arXiv: 1310.1531 (2013).
- VI Ezgii Erturk, Ebru Akapinar Sezer "A comparison of some soft computing methods for software fault prediction" Expert system with applications, Elsevier, pp1872-1879, vol.42, 2015.
- VII F. Al-Shargie, T.B. Tang, N. Badruddin, & M. Kiguchi, (2018). Towards multilevel mental stress assessment using SVM with ECOC: an EEG approach. Medical & biological engineering & computing, 56(1), 125-136.
- VIII Fruits 360 Dataset on Kaggle. <https://www.kaggle.com/moltean/fruits>. last visited on 06.07.2019
- IX G. Muhammad, (2015). Date fruits classification using texture descriptors and shape-size features. Engineering Applications of Artificial Intelligence, 37, 361-367.
- X H. M. Zawbaa, M. Abbass, M. Hazman, & A. E. Hassenian, (2014, November). Automatic fruit image recognition system based on shape and color features. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 278-290). Springer, Cham.
- XI L. Qiang, C. Jianrong, L. Bin, D. Lie, & Z. Yajing, (2014). Identification of fruit and branch in natural scenes for citrus harvesting robot using machine vision and support vector machine. International Journal of Agricultural and Biological Engineering, 7(2), 115-121.
- XII M. Achirul Nanda, K. Boro Seminar, D. Nandika, & A. Maddu, (2018). A comparison study of kernel functions in the support vector machine and its application for termite detection. Information, 9(1), 5.
- XIII P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," in Encyclopedia of Database Systems, L. LIU and M. T. ÖZSU, Eds. Boston, MA: Springer US, 2009, pp. 532–538.
- XIV R. S. Chora's, "Image Feature Extraction Techniques and their Applications for CBIR and Biometrics Systems". International Journal of Biology and Biomedical Engineering, 2007, 1(1), 6–16.
- XV S. Fatima, and M. Sesehashayee, (2020). Healthy Fruits Image Label Categorization through Color Shape and Texture Features Based on Machine Learning Algorithm, International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075, Volume-9 Issue-3.
- XVI S. R. Dubey, & A. S. (2012). Robust approach for fruit and vegetable classification. Procedia Engineering, 38, 3449-3453.

- XVII S. M. Iqbal, A. Gopal, P. E., Sankaranarayanan, & A. B. Nair, (2016). Classification of selected citrus fruits based on color using machine vision system. International journal of food properties, 19 (2), 272-288.
- XVIII Shepperd, D. Bowes, and T. Hall, "Researcher bias: The use of machine learning in software defect prediction," Software Engineering, IEEE Transactions on, vol. 40, no. 6, pp. 603-616, 2014.
- XIX S. Ibrahim, N. A. Zulkifli, N. Sabri, A. A. Shari, & M. R. M. Noordin, (2019). Rice grain classification using multi-class support vector machine (SVM). IAES International Journal of Artificial Intelligence, 8 (3), 215.
- XX T. G. Dietterich, G. Bakiri, 1995. Solving Multiclass Learning Problems via Error-Correcting Output Codes. Journal of Artificial Intelligence Research, 2, pp. 263-286.
- XXI Y. Zhang, & L. Wu, (2012). Classification of fruits using computer vision and a multiclass support vector machine. Sensors, 12 (9), 12489-12505.
- XXII Z. Wen, B. Li, R. Kotagiri, J. Chen, Y. Chen, & R. Zhang, (2017, February). Improving efficiency of SVM k-fold cross-validation by alpha seeding. In Thirty-First AAAI Conference on Artificial Intelligence.
- XXIII Z. Yan, Y. Yang, (2014). Application of ecoc svms in remote sensing image classification. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40 (2), 191.
- XXIV Z. Yan, & Y. Yang, (2014). Performance analysis and coding strategy of ECOC SVMs. International Journal of Grid and Distributed Computing, 7 (1), 67-76