Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2021.020405
ISSN 2790-0983 Vol. 2, Issue 4: 34-41

Maize Variety Identification Based on Support Vector Machines

Romany Jalali*

Dhurakij Pundit University, Thailand

**corresponding author*

Keywords: Support Vector Machine, Maize Varieties, Variety Identification, Feature Parameters

Abstract: Seeds are the most basic and critical means of production in agricultural production. Declining purity of seed varieties will significantly reduce crop yield and product quality. The quality of crop seed varieties, especially purity testing, can protect the interests of farmers and promote the development of rural economies. The aim of this paper is to study maize variety identification based on support vector machines. Based on the morphological structure characteristics of maize varieties, a set of characteristic parameters that can accurately reflect the morphological structure of different maize varieties is proposed for variety identification. The parameter set consists of four components: shape, colour, size and sharpness parameters. The experimental results show that the support vector machine algorithm outperforms the BP neural network.

1. Introduction

As one of the world's top three crops, maize varieties have an important influence on the cultivation and use of maize, especially in the cultivation process, where the identification of maize varieties is particularly important, and the yield and quality of different varieties of maize seed are also closely related to maize yield and quality. Seed quality is directly related to the quality of crop production and product quality [1-2]. With the improvement of people's living standards and the development of the agricultural economy, the market demand for agricultural products is not only about quantity, but also increasingly about variety quality [3].

With the development of computer technology, digital image processing and pattern recognition, many experts and scholars have conducted extensive research on food variety identification [4]. Rabih Nachar described a system for automatic identification of food products. Color and thermal images are received by an electronic processor. At least one feature is extracted from the recognized pixel area in the color image, and the food type corresponding to the food is automatically recognized based on the extracted feature. [5].R. Manoranjitham classifies the satellite image map by a new model using a convolutional neural network modified by a support vector machine. The dataset is from Kaggle and uses satellite images of the Amazon rainforest to train a multi-label classifier. By training the model proposed by CNMSVM, an F2 score of at least 0.91412 was achieved [6]. Athanasios Psaltis developed a handwritten alphanumeric string recognition system based on an embedded development platform. The S3C2440 development board was chosen for the hardware platform. The image processing uses OpenCV and support vector machine (SVM) classification techniques to achieve handwritten alphabet and number recognition. The system has the advantages of low power consumption and low memory consumption [7]. In conclusion, the automation of the identification of different maize varieties is of great importance in reducing costs, saving human and material resources and improving the efficiency of agricultural experts, and is also a future trend in the development of agricultural automation in China, which has important theoretical and practical significance for the development of intelligent agriculture in China, and has positive and far-reaching significance for accelerating the pace of China's agricultural economy.

Maize variety identification is the study of the authenticity and purity of maize seed varieties. This paper identifies the problems to be solved in the field of maize quality inspection in China and conducts research on maize variety identification based on maize seed morphology on the basis of digital image processing technology and related knowledge. Based on the characteristics of the tip position of maize seeds, i.e. the tip position of normal maize seeds is the position with the greatest curvature of the maize seed boundary, an algorithm for quickly finding the tip position of maize seeds is proposed with high accuracy making the calculation of characteristic parameters of maize seeds highly reliable. It is important to accelerate the scientific and technological progress of the agricultural products processing industry, promote the quality of agricultural products, improve the international competitiveness of Chinese agricultural products, regulate the market, high quality and optimal price, and improve farmers' income.

2. A Study of Maize Variety Identification Based on Support Vector Machines

2.1. Support Vector Machine Algorithms

Support vector machines are based on statistical learning methods and use the principle of structural risk minimisation with good generalisation capabilities [8-9]. When linearly separable, there is a hyperplane that completely separates the training set, which can be described as:

$$w \bullet x + b = 0 \tag{1}$$

Where " ·is the dot product, w is an n-dimensional vector and b is the offset.

In equation (2) its solution must satisfy:

$$\alpha_i [y_i (w * x + b) - 1] = 0. i = 1,2,3,...n \tag{2}$$

Those samples with αi = 0 have no effect on the classification, and only those samples with αi > 0 play a role in the classification, and these samples are called support vectors, and the expression of the classification indicator function is obtained as :

$$f(x) = \sum_{i=1}^{n} y_i \alpha_i (x^* x_i) + b^* \tag{3}$$

The sign of f(x) is used to determine the attribution of x.

There are two approaches in SVM to solve multi-categorization problems, the first is a one-to-one approach and the second is a one-to-many approach. The one-to-many strategy focuses on one category corresponding to all other remaining categories [10-11]. For example, if there are six categories of maize varieties, then six SVM classifiers are trained correspondingly. For the first classifier category 1 is one category, the other five categories are one category, for the second classifier category 2 is one category, the remaining five categories are one category and so on for a total of n categories of classifiers trained. The one-to-one strategy is to train a classifier between each category, so if there are n categories then a total of n(n-1)/2 classifiers have to be trained. When an unknown sample comes in, each classifier is used to test it, and finally the final category is decided using the voting principle [12-13].

2.2. Maize Variety Image Processing

(1) Threshold segmentation based on the Otsu method

In this study, in order to facilitate the detection of maize tips and feature extraction later, the enhanced image needs to be segmented. Due to the non-uniform grey value of the background, only part of the maize seeds can be segmented with one-dimensional or two-dimensional thresholding, most of which will be segmented into the background colour, and if multiple thresholding is used, each threshold needs to be calculated, which is complicated to calculate and also cannot guarantee the correctness of the threshold selection, therefore In this paper, the adaptive thresholding Otsu algorithm is used to segment the image of maize kernels [14-15].

The total mean grey level of the image is denoted as u and the variance between classes is denoted as g. Then we have:

$$u = u_0 \times w_0 + u_1 \times w_1$$
$$g = w_0 \times (u - u_0)^2 + w_1 \times (u - u_1)^2 \tag{4}$$

The obtained t is the optimal segmentation threshold when the variance g between classes is maximum [16].

(2) Harris algorithm based maize seed tip detection

In the Harris algorithm, the magnitude of the response value R indicates the degree of sharpness of the point. From the images of seeds with tip features of different maize varieties, the tip is the most obvious angular feature and is usually sharper than other features [17-18]. The algorithm for maize seed tip recognition in this paper is as follows:

Input:Binary image of maize seeds I

(1) Calculate the gradient of the binary image

(2) Calculate the covariance matrix of the gradient image

(3) Smooth the covariance C with the w(x, y) weighting function;

(4) Calculate the angular response of the image;

(5) The point with the largest response value is considered to be the tip of the corn kernel.

3. A Survey and Study of Maize Variety Identification Based on Support Vector Machines

3.1. LIBSVM Software

There are many software packages related to SVM computation such as LIBSVM, mySVM etc. LIBSVM is a simple, easy to use, fast and effective general purpose SVM package for solving classification, regression and other problems. The software source code is open for improvement, modification and application by other operating systems. Therefore, the study decided to use the software as the working software.

3.2. Characteristic Parameters

Morphological and structural characteristics of maize seeds, including colour, size, shape and canopy characteristics. The colour characteristics are the colour of the surface of the maize kernel and the colour of the endosperm. Size characteristics are the size of the maize kernels. Shape characteristics are the shape of the maize kernel. Canopy characteristics are the characteristics of the corn seeds that protrude from the top. There are two types: smooth and non-smooth. In the actual automatic seed quality inspection line, the high speed movement and position of the seeds are also random and it is difficult to detect the crown features of the seeds, therefore, in this study, the crown features of the maize seeds were not detected and only the sharpness of the tips of the maize seeds was detected. The flow chart for calculating the feature parameters in this study is shown in Figure 1:

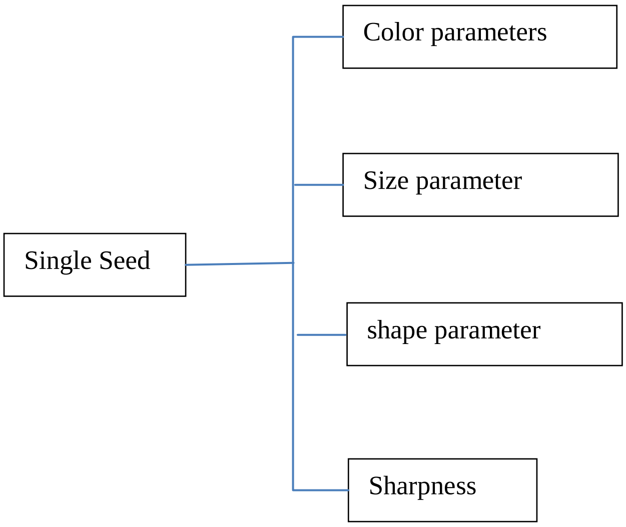

Figure 1. Calculation diagram of seed characteristic parameters

3.3. Classification Methods

In this paper, a directed acyclic graph organisation classifier is built. When classifying, we can first ask the classifier "1 vs 5" (i.e. we can answer "1 or 5 classes"). If the answer is "5", we ask the classifier on the left "2 vs 5". If it says "5", we can move on to the 5 side.

For the 1-V-1 category, when the number of categories is 5 (if the number of categories is k and the classifier is k (k-1)/2), 10 classifiers should be called. Instead of using the classifiers we created, we only call four classifiers (or k-1 classifiers if the classifier number is k). These classifiers can be classified quickly and without mismatches, and cannot be classified as phenomena.

4. Analysis and Research of Maize Variety Identification Based on Support Vector Machines

4.1. Classification Results

In this paper, the LIBSVM 2.83 package easy.py was used for parameter preference and variety identification of the model, aided by the installation of python and the plotting tool gnuplot. Five maize seed varieties of 200 kernels were used as training samples, 40 kernels of each maize. A recognition sample of 250 seeds, 50 of each maize, was used. Radial basis parameters were selected via the LIBSVM 2.83 package grid.py with optimal parameters C=30 and g=0.008821. The results of the five maize seed variety identification are shown in Table 1.

Table 1. Recognition results of corn seed varieties by support vector machine

Sample name	Fine jade 99	Maidenhead 6702	Tunyu 808	Dragon height L2	Dredging order 29
Training samples (pcs.)	40	40	40	40	40
Test sample (pcs.)	50	50	50	50	50
Recognition accuracy (%)	98	96	97	97	99

4.2. Comparison of Algorithms

In this paper, a BP network recognition system was constructed through the artificial neural network toolbox of MATLAB 7.0, as shown in Figure 2, to conduct training and recognition experiments on five maize seed varieties, with the following steps:

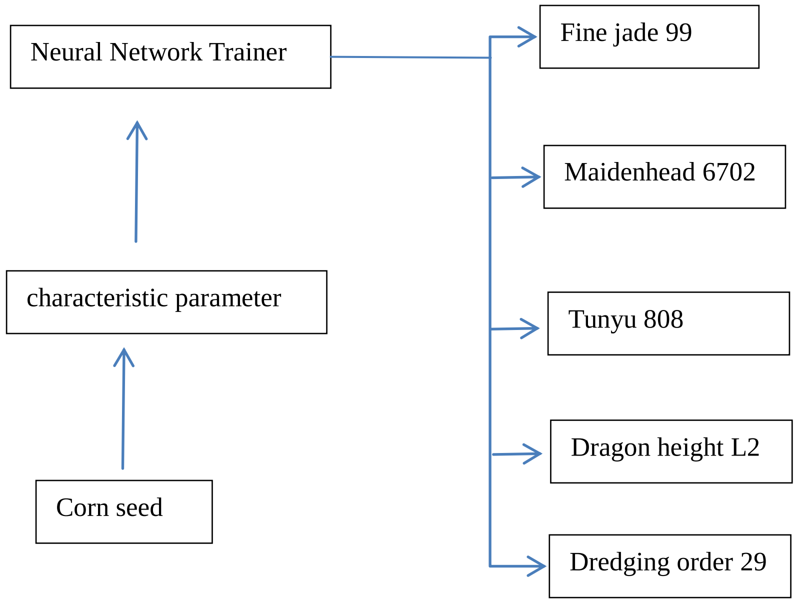

Figure 2. BP network identification system

Different network structures have an impact on the accuracy of recognition. In this paper, a BP network with a network model of $5 \times n \times 5$ was constructed to train and recognize the samples. Through Table 2, it can be obtained that the recognition accuracy is highest when n=9.

Table 2. Identification results of different network structures

Network structure input - hidden layer - output	Accurate identification of test samples (%)
$5 \times 5 \times 5$	93
$5 \times 6 \times 5$	92
$5 \times 7 \times 5$	92
$5 \times 8 \times 5$	91
$5 \times 9 \times 5$	94
$5 \times 10 \times 5$	92

A training sample of 200 kernels of each of the five maize seed varieties was selected, 40 kernels of each maize. The recognition sample was 250 kernels, 50 kernels of each type of maize. The recognition results are shown in Figure 3.

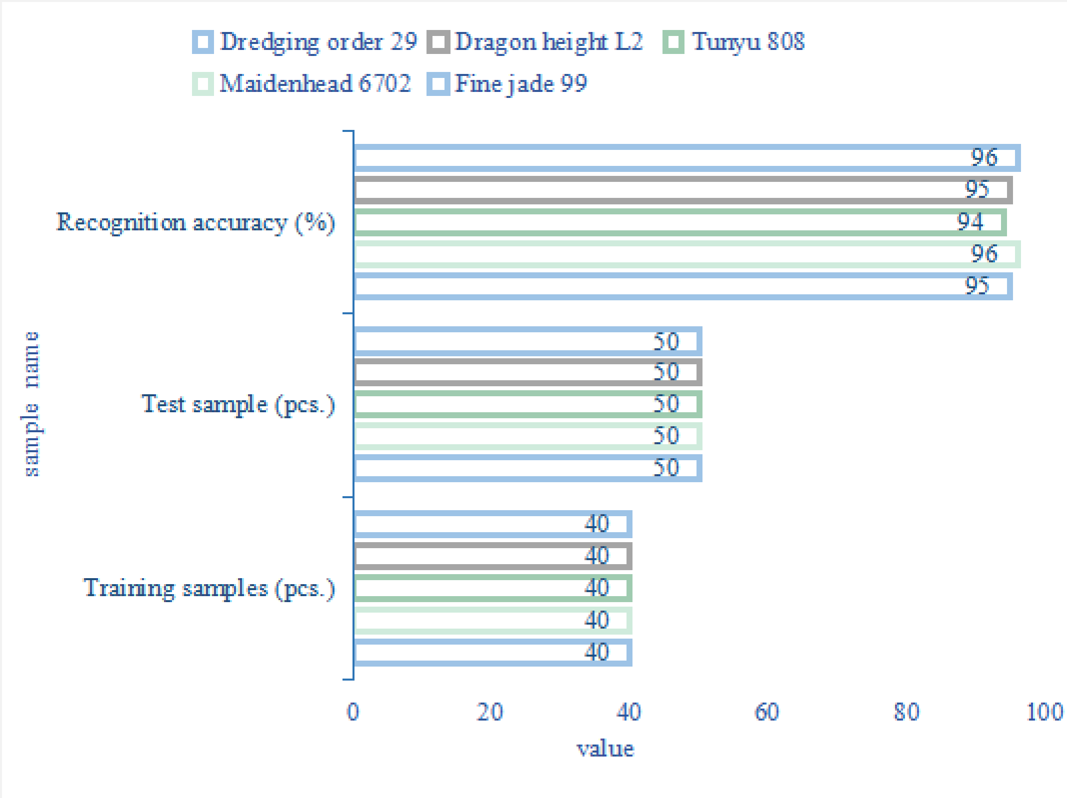

Figure 3. Recognition results of corn varieties based on BP neural network

Compared with Table 1 and Figure 3, the average recognition rate of SVM is higher than the average recognition rate of BP neural network. This is a good example of the superiority of Support Vector Machines, a new automatic learning algorithm based on statistical learning theory for small sample and predictive learning. Although it can improve the recognition rate of even the most complex artificial neural network models, it can easily form local minima in the absence of global optimisation, increased training times, low learning efficiency and slow convergence.SVM is a new sample learning method based on robust theory. It essentially does not involve probability measures or the law of large numbers and is therefore different from existing statistical methods. It essentially avoids the traditional inductive inference process and enables efficient "transfer inference" from training samples to prediction samples, greatly simplifying common classification and regression problems.

5.Conclusion

This paper presents some research on the identification of maize varieties, the results of which can provide the basis for rapid and non-destructive testing of maize variety identification in the future. Of course, the methods studied in this paper are not yet mature enough for practical application and there are still some aspects waiting to be improved and developed. At the end of the paper, through analysis and research, there are several suggestions for the shortcomings and areas for further research on the use of computer vision for maize variety identification:When using images of maize seeds collected by digital cameras in unenclosed environments, several conditions are limited in this paper due to differences in the environment and imaging conditions. It is worthwhile to investigate further how to gradually reduce the limiting conditions. When extracting feature parameters, adding a few more conventional parameters may result in a higher recognition rate of maize varieties. Due to limited effort and capacity, this paper has not been studied in depth.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Helena Gómez-Adorno, Roddy Fuentes-Alba, Ilia Markov, Grigori Sidorov, Alexander F. Gelbukh: A convolutional neural network approach for gender and language variety identification. J. Intell. Fuzzy Syst. 36(5): 4845-4855 (2019) https://doi.org/10.3233/JIFS-179032

[2] Silvia Cascianelli, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara: Boosting modern and historical handwritten text recognition with deformable convolutions. Int. J. Document Anal. Recognit. 25(3): 207-217 (2021)

[3] Panagiotis Tsinganos, Bruno Cornelis, Jan Cornelis, Bart Jansen, Athanassios Skodras: Hilbert sEMG data scanning for hand gesture recognition based on deep learning. Neural Comput. Appl. 33(7): 2645-2666 (2020) https://doi.org/10.1007/s00521-020-05128-7

[4] Monika Bansal, Munish Kumar, Manish Kumar, Krishan Kumar: An efficient technique for object recognition using Shi-Tomasi corner detection algorithm. Soft Comput. 25(6): 4423-4432 (2020) https://doi.org/10.1007/s00500-020-05453-y

[5] Rabih Nachar, Elie Inaty, Patrick J. Bonnin, Yasser Alayli: Hybrid minutiae and edge corners feature points for increased fingerprint recognition performance. Pattern Anal. Appl. 23(1): 213-224 (2018) https://doi.org/10.1007/s10044-018-00766-z

[6] R. Manoranjitham, P. Deepa: Efficient invariant interest point detector using Bilateral-Harris corner detector for object recognition application. Multim. Tools Appl. 77(8): 9365-9378 (2017) https://doi.org/10.1007/s11042-017-4982-5

[7] Athanasios Psaltis, Konstantinos C. Apostolakis, Kosmas Dimitropoulos, Petros Daras: Multimodal Student Engagement Recognition in Prosocial Games. IEEE Trans. Games 10(3): 292-303 (2017) https://doi.org/10.1109/TCIAIG.2017.2743341

[8] Aamir Rasool, Abdul Rehman Javed, Zunera Jalil: SHA-AMD: sample-efficient hyper-tuned approach for detection and identification of Android malware family and category. Int. J. Ad Hoc Ubiquitous Comput. 38(1/2/3): 172-183 (2021) https://doi.org/10.1504/IJAHUC.2021.10042892

[9] Hossein Moosaei, Fatemeh Bazikar, Saeed Ketabchi, Milan Hladík: Universum parametric-margin ν-support vector machine for classification using the difference of convex functions algorithm. Appl. Intell. 52(3): 2634-2654 (2021) https://doi.org/10.1007/s10489-021-02402-6

[10] Hossein Moosaei, Fatemeh Bazikar, Saeed Ketabchi, Milan Hladík: Universum parametric-margin ν-support vector machine for classification using the difference of convex functions algorithm. Appl. Intell. 52(3): 2634-2654 (2021) https://doi.org/10.1007/s10489-021-02402-6

[11] Ashish Sharma: Stochastic nonparallel hyperplane support vector machine for binary classification problems and no-free-lunch theorems. Evol. Intell. 15(1): 215-234 (2021) https://doi.org/10.1007/s12065-020-00503-8

[12] Satoshi Kodama: Shape classification based on solid angles by a support vector machine. Intell. Data Anal. 26(4): 933-957 (2021) https://doi.org/10.3233/IDA-215992

[13] M. Kusuma Sri, E. Gomathi: Classification of breast cancer images using completed local ternary pattern and support vector machine. Int. J. Bioinform. Res. Appl. 18(1/2): 130-140 (2021) https://doi.org/10.1504/IJBRA.2021.10045876

[14] Talal Almutiri, Faisal Saeed: A hybrid feature selection method combining Gini index and support vector machine with recursive feature elimination for gene expression classification. Int. J. Data Min. Model. Manag. 14(1): 41-62 (2021) https://doi.org/10.1504/IJDMMM.2021.122038

[15] Krishnamoorthy Sashi Rekha, Suthanthira Amalraj Miruna Joe Amali: Efficient feature subset selection and classification using levy flight-based cuckoo search optimization with parallel support vector machine for the breast cancer data. Int. J. Imaging Syst. Technol. 32(3): 869-881 (2021) https://doi.org/10.1002/ima.22686

[16] Mohammed Alshutbi, Zhiyong Li, Moath Alrifaey, Masoud Ahmadipour, Muhammad Murtadha Othman: A hybrid classifier based on support vector machine and Jaya algorithm for breast cancer classification. Neural Comput. Appl. 34(19): 16669-16681 (2021)

[17] Sana Ullah Khan, Naveed Islam, Zahoor Jan, Khalid Haseeb, Syed Inayat Ali Shah, Muhammad Hanif: A machine learning-based approach for the segmentation and classification of malignant cells in breast cytology images using gray level co-occurrence matrix (GLCM) and support vector machine (SVM). Neural Comput. Appl. 34(11): 8365-8372 (2021) https://doi.org/10.1007/s00521-021-05697-1

[18] Taisei Nohara, Satoshi Oyama, Itsuki Noda: Pairwise classification using quantum support vector machine with Kronecker kernel. Quantum Mach. Intell. 4(2): 1-13 (2021)