Identification of pests hidden in wheat kernels based on support vector machine classifier

Zhihui Li[1,*], Tong Zhen[1],Yuhua Zhu[2]

1. College of Information Science and Engineering, Henan University of Technology Zhengzhou 450001, China

2. Yellow River Conservancy Technical Institute, Kaifeng 475000, China

zhihui_li511@sina.com

ABSTRACT. *The identification of pests hidden in stored wheats, essential to grain storage safety, is a key difficulty in the research of target detection. This paper introduces the support vector machine (SVM) classifier to identify the pests hidden in wheat kernels, and selects the proper kernel function and parameters to classify various samples. It is verified that the proposed method could accurately detect the pests in wheat kernels. This research provides new insights into the application of pattern recognition in bio-photon detection of pests in stored grains.*

RÉSUMÉ. *L'identification des parasites cachés dans les blés stockés, essentielle à la sécurité du stockage du grain, est une difficulté majeure dans la recherche sur la détection des cibles. Cet article présente le classifieur de machine à vecteurs de support (en anglais support vector machine, SVM) pour identifier les parasites cachés dans les noyaux de blé et sélectionne la fonction et les paramètres du noyau appropriés pour classifier divers échantillons. Il est vérifié que la méthode proposée pourrait détecter avec précision les parasites dans les noyaux de blé. Cette recherche fournit de nouvelles perspectives sur l'application de la reconnaissance de formes à la détection par bio-photon des parasites cacshés dans les grains stockés.*

KEYWORDS: *grain kernels, support vector machine, classification, characteristic parameter.*

MOTS-CLÉS: *noyaux de céréales, machine à vecteurs de support, classification, paramètre caractéristique.*

1. Introduction

Harmful biological hazard is a serious problem to the safe storage of grain, so accurately detecting and positioning pests has become an important task to maintain grain safety. The current research methods only focus on the research of granary pests' adults stage. Temperature and moisture are the main factors that determine the quality of grain storage. Excessively high or low temperature and moisture can cause some adverse reactions, such as the growth of pests, bacteria and fungi. So the quantity of pests can indirectly reflect the temperature and humidity. However, the detection of stored grain pests is a difficult problem (Liang *et al*., 2014; Wang *et al*., 2014).

Granary generally has a tight, solid civil structure, less affected by external climate factors, warehouse temperature, grain temperature is relatively stable, food nutrition is sufficient, living environment is stable. For a long period of time and low mortality, the occurrence of pests has rapid growth population in the short term to cause great harm. Initial pests can directly endanger the whole grain concealing in the grain. Such as corn elephant, rice elephant, etc. In the complete grain kernels, the initial pests usually occur first, then the late pests will occur. The emergence of late-stage insect pests indicates that the grain hazards have been serious to a certain extent. The current grain storage information detection technology can completely detect the species and density of late-stage pests or adults, but there is no clear detection method for pests hidden in grain kernels.

Referring to the application of intelligent technology (Bello and Mamman, 2018) and algorithm in detection technology (Abdullahi *et al*., 2018), this paper uses biophoton technology to detect hidden pests in grain kernels. Based on the theory of quantum optics, the microscopic mechanism and physical essence of photon-life interaction are studied by using the weak light and ultra-weak photon radiation detection system of life system as experimental means, and the photophysical essence of biological tissues and organisms.

In view of the limitations and shortcomings of the existing hidden grain pest detection methods, in this paper, the bio-photon analysis technology and pattern recognition technology are introduced. The SVM method is used to design the classifier. The kernel function is determined and compared with the classification and recognition results by using different feature vectors. The research in this paper has laid a foundation for the hidden pest bio-photon detection model.

This paper first introduces the experimental method, then introduces the SVM algorithm and experiment design. Finally, the experimental results are analyzed and summarized.

2. Signal acquisition

2.1. Experimental method

Full-particle, uniform wheat kernels were selected for the experiment. The wheat kernels were washed three times with pure water and dried until the moisture was 14% and then put into a sealed bag for later use. The temperature of the drying box is 60 ℃, which could kill all the pests in wheat. Maize weevils were cultivated and spawned as the grain pest, at an optimum moisture content of 12.5% to 14%.

200 energetic maize weevils and 200g of normal wheat were put into the feeding bottles. The adults were screened out and two days later, the wheat was infected with hidden pests in the feeding bottles. The feeding bottle was put into the biochemical incubator. Another 200g of normal wheat was selected without any treatment as the control samples. The infected wheat and normal wheat were divided into 10 uniform feeding bottles and the bio-photon of each 3 groups were measured. The temperature in the whole measurement and storage process was set at 28 ℃, which is the optimum culture temperature.

2.2. Parameter extraction

Characteristic parameters like mean, mode, median, interquartile range and mean deviation, variance, standard deviation, coefficient, skewness (Cortes and Vapnik, 1995; Thilina *et al*., 2012), kurtosis and photon statistical entropy are used for classification. Mode is the value that appears most often in a set of data. If we sort the data from large to small, the median is the value located in the middle of the sequence. Data distributed from the 25th to 75th percentile are within the interquartile range.

The mean of data in the type definition see Eq. (1):

$$\bar{x} = 1/N\sum x_i \,(i=1, \ldots, N) \tag{1}$$

MD is the mean deviation. It can be obtained from the below calculation Eq. (2):

$$MD = \sum |x - \bar{x}|/N \tag{2}$$

Variance is a parameter describing the dispersion degree of a dataset or random variables. It is used to calculate the difference between each observed variable and the mean. In practical work, it is difficult to obtain the population mean, so it is replaced by the sample statistic. After correction, the variance of the sample is calculated as follow Eq. (3):

$$S^2 = 1/(N-1)\sum (x_i - \bar{x})^2 \,(i=1, \ldots, N) \tag{3}$$

Standard deviation coefficient is the ratio of the standard deviation and the mean, which is used to indicate the relative index of the data variation degree, see Eq. (4).

$$V_s = 100\% \times \sqrt{S^2}/|X| \tag{4}$$

Skewness is used to describe the probability distribution asymmetry of real valued random variables. The calculation formula is shown in Eq. (5):

$$SK = (n\sum(x_i - \bar{x})^3)/(n-1)(n-2)sd^3 \tag{5}$$

Kurtosis is used to characterize the peak height of the probability density distribution curve around the mean. The calculation formula is shown in Eq. (6):

$$kurtosis = (\sum(x_i - x)^4)/(N-1)S^4 \tag{6}$$

Photon statistical entropy is mainly used to characterize the photon number distribution deterministic value which is put forward based on ultra-weak luminescence photon distribution that can be measured directly. The calculation formula is shown in Eq. (7):

$$G = -\sum P(n)\ln P(n) \ (n=0, \ldots, \infty) \tag{7}$$

G is the statistical entropy of photon and *P(n)* is the photon number distribution.

3. Problems of pattern classification based on SVM

3.1. SVM basic principle

In machine learning, SVM can be used for pattern recognition, data analysis, classification and regression analysis, generally speaking, it has a lot of advantages in solving small-sample (Alamdar *et al*., 2016), nonlinear and high dimensional pattern recognition. It is a two-class classification model, whose basic model is a linear classifier with maximum interval defined in feature space (Li *et al*., 2015; Chou *et al*., 2014; Van *et al*., 2013). The learning strategy of SVM is distance maximization, which finally transforms into solving a convex quadratic programming problem. The two-class sample classification is shown in Figure 1 (Devos *et al*., 2014; Feng and Palomar, 2015; Maldonado *et al*., 2014).
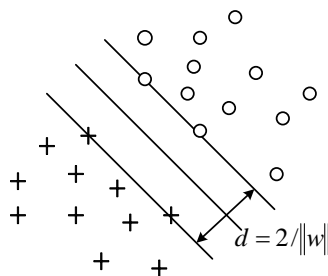

Figure 1. Schematic diagram of the optimal classification hyper plane

The theory of SVM is that, assuming a given training sample set is $\{(x_i, y_i), i=1, ..., m, x_i \in R\}$, the expected output is $y_i \in \{-1, +1\}$, where -1 and +1 respectively represent two linear types that divide the samples into two parts. To seek a class similar to this distribution is independent of the new samples for the correct division of the optimal classification hyper plane which is called classification decision, where the linear discriminant function is g(x)=wx+b. Two types of sample classification diagrams are shown in Figure 1, where the circles and plus represent two kinds of samples. SVM is used to find the optimal classification plane in the final result. It can not only identify different types of samples, but also maximize the classification interval. When SVM is applied to solve nonlinear regression and time series prediction problems, it is usually called SVR (support vector machine).

The equation of hyper plane is wx+b=0. It is normalized to make all samples satisfy the equation $|g(x)| \geq 1$. It means that when the distance classification of the recent samples meets $|g(x)|=1$, the classification interval is equal to $2/\|w\|$. If the $\|w\|$ is minimum, the surface classification of all the samples are classified correctly, See Eq. (8):

$$y_i(wx_i+b)-1 \geq 0, i=1, ..., m \tag{8}$$

To solve the constrained optimization problem, the extended Lagrange multiplier theory is applied to construct the objective function $1/2\|w\|$ minus Lagrangian by constraint function $a_i > 0$ and sub product as the Lagrange function, which can be written as follow Eq. (9):

$$L(w, b, a)=1/2\|w\|-a[y(wx+b)-1] \tag{9}$$

The solution to the constrained optimization problem is obtained from the saddle point of the Lagrange function, which satisfies the partial derivative of W and B at the saddle point to zero. It is the dual problem of the primal problem maxL(w, b, a). s.t.$(\partial L(w, b, a))/\partial w = w - \sum a_i y_i x_i = 0$, $(\partial L(w, b, a))/\partial b = \sum a_i y_i = 0$.

We can obtain the optimal solution $a*i$ and the best classification distance $w* = \sum y_i a* ix_i$, and also the optimum bias $b* = y_j - \sum y_i a* i(x_i \cdot x_j)$, according to the aforementioned formula. We can use one data as $a_i \neq 0$ to obtain the value of w* after determining the w, and finally obtain the optimal classification boundary. The sample data are also known as support vectors when a_i is not equal to zero. The optimal hyper plane is a linear combination of support vectors (Maldonado *et al.*, 2014; Nalepa *et al.*, 2015; Sun *et al.*, 2015). The discriminant function of SVM is shown in Formula, where m denotes the set of support vectors, see Eq. (10):

$$f(x)=sgn(\sum_{i \in M} a* iy_i <x_i, x>+b*) \tag{10}$$

SVM can map the sample data which cannot be separated linearly to high-dimensional feature space (Hilbert space) through nonlinear mapping $\varphi(x)$ by introducing the kernel function, and then do linear processing in high dimensional feature space. K(x, y) is a valid kernel function when it satisfies the Mercer's condition as $K(x, y)=\varphi(x)\cdot\varphi(y)$, where $\varphi(x)$ is a mapping expression. The

classification function in high-dimension space can be obtained when a proper kernel function is chosen, which is shown in Eq. (11):

$$f(x)=sgn(w{\cdot}\phi(x)+b)=sgn(\textstyle\sum_i a_i y_i \phi(x_i){\cdot}\phi(x)+b) \tag{11}$$

The support vector machine has avoided the dimension disaster with the inner product operation. The learning machine is constructed to implement the input space in different types of nonlinear decision surfaces, resulting in different SVM algorithms. The commonly used kernel functions and its formula are as follows:

(1) Linear Kernel

The linear kernel is used mainly for linear separation. When the dimension in feature space is the same as in the input space, the classification effect is desirable for linear separable data with few parameters and fast convergence. So the linear kernel function is usually used first for classification. PCA transforms the original data into a set of linear independent representations in each dimension by linear transformation, which can be used to extract the main feature components of the data. KPCA is often used for dimension reduction. The formula is shown in Eq. (12) (Zhang *et al*., 2014):

$$k(x,\ y)=x^T y+c \tag{12}$$

(2) Polynomial kernel

Polynomial kernel function can map a low dimensional feature space to a high dimension, but it has many parameters. When the order of polynomial is relatively high, the matrix element value will be infinite or infinitesimal, so the computational complexity will be added. The formula is shown in Eq. (13):

$$k(x,\ y)=(ax^T y+c)^d \tag{13}$$

In this formula, the constant term c, the slope α and the polynomial degree d are adjustable.

(3) Gaussian kernel

The Gaussian radial basis function is a strong local kernel function that can map a sample to a higher dimensional space. It is one of the most widely used functions. Whether the sample size is big or small, it will have a good performance and less parameters compared with the polynomial kernel function. Therefore, in most cases, the Gaussian kernel function is a preferred choice, see Eq. (14):

$$k(x,\ y)=exp(-(\|x-y\|^2)/2\sigma^2) \tag{14}$$

Alternatively, the Gaussian kernel can also be implemented using, see Eq. (15):

$$k(x,\ y)=exp(-\gamma\|x-y\|^2) \tag{15}$$

In the above formula, the value of sigma is important to the performance of the kernel function. It cannot be overestimated or underestimated.

(4) Hyperbolic Tangent (Sigmoid) Kernel

When the sigmoid kernel function is used, the support vector machine is a multi-layer neural network. Therefore, during the selection of kernel function, if there are some prior knowledge about the data, they can be used to select the kernel distribution; if not, cross validation is used commonly. The error rate is the best indicator for different kernel functions. Multiple kernel functions can be integrated together to form a mixed kernel function.

Firstly it comes from the neural network field, so it is also known as the Multilayer Perceptron (MLP) kernel and the bipolar sigmoid function is often used as an activation function for artificial neurons, see Eq. (16)

$$k(x, y)=\tanh(ax^Ty+c) \tag{16}$$

Although only conditionally positive, it performs well in practical applications. There are two adjustable parameters in the K function, the slope and the intercept constant C. Alpha is usually valued at 1/N (N is the data dimension).

3.2. Design process of SVM

SVM has a strong mathematical theory basis, which has been put forward specifically for the small sample learning problem. The two programs are used to find the optimal solution in theory, so the global optimal solution can be obtained to solve the problem of local minimum in the neural network. Through nonlinear mapping to high dimensional feature space, the linearly inseparable pattern of the low dimensional space can be linearly separable, but if the classification or regression is carried out in high dimensional space, there will be problems with the nonlinear mapping function form, parameters and feature space dimension, and the biggest obstacle of all is that there exists a dimension disaster in the high dimensional feature space. The kernel function can solve the nonlinear problem in this problem effectively. In this paper, SVM is used to identify wheat kernels with hidden pests in order to classify wheat kernels with hidden pests and normal kernels. The time domain characteristics of the test set are used as the input parameters of SVM.

This research mainly uses the SVM toolbox of Matlab to carry out the classifier design. The classification decision flowchart is shown in Figure 2.

When SVM is used for classification, the kernel function of SVM is determined according to the characteristics of the input feature vectors. Various types of kernel functions can be constructed in different feature spaces, which can produce different decision surfaces, i.e. the SVM algorithms. The kernel function is selected according to the feature dimension and the input feature vector. Table 1 shows the classification accuracies of different kernel functions under different kernel functions. The time domain characteristic parameters of the spontaneous emission of the wheat kernels are used as the input feature vectors of SVM. In the classification process, the global variables (which define different kernel function parameters) are

chosen from small to large in order to achieve the best classification results. Evaluation parameters such as classification accuracy, detection rate and false alarm rate are used to evaluate the effect.
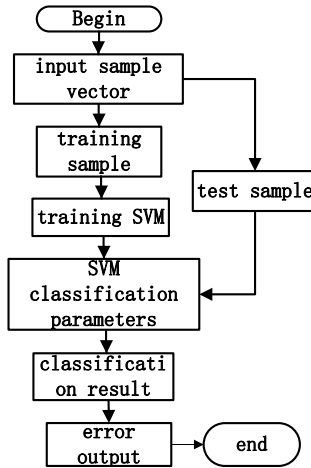

Figure 2. Flow chart of SVM classification

Table 1. Classification results of SVM under different kernel functions

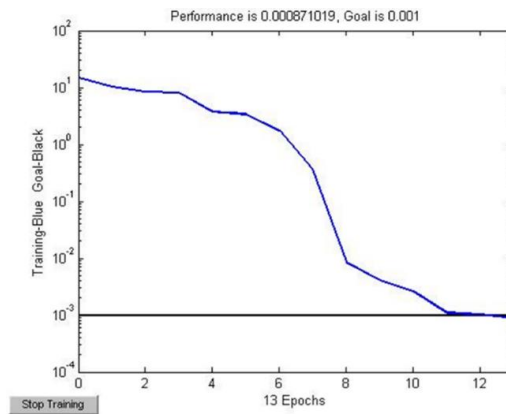| Kernel type	linear	poly	Gaussian	sigmoid
The missing rate	26.89%	20.00%	6.67%	23.34%
False alarm rate	33.56%	20.00%	13.33%	23.34%

Figure 3. The error curve when used Gaussian as a kernel function

By comparing the classification results of different kernel functions, we can find that the Gaussian kernel function can achieve the highest classification accuracy and the lowest missing rate and false alarm rate. So the Gaussian function was selected as the kernel function to classify the spontaneous emission signals of wheat in this experiment. The training error met the set requirements after roughly 12 cycles when the Gaussian kernel function was used to do the classification, and the error curve in the network error training process is shown in Figure 3.

3.3. Training and testing of SVM

In this paper, support vector machine is used for discriminant classification. The data of the test group and the control group were processed before the experiment. The time domain and frequency domain feature vectors are extracted and training and test samples were selected.

The training samples were used to train the SVM, and then the SVM discriminant function was used to classify the test samples. The classification is output with various indicators with error function. The corresponding classification results are provided when the input of different characteristic parameters is given.

The time domain characteristic parameters of the three groups of samples obtained from the same test conditions were classified by the SVM classifier. The classification results are listed in Table 2.

Table 2. SVM classification results of three groups of experimental data in time domain

Evaluating indicator	Group 1	Group 2	Group 3
Recognition accuracy	86.67%	90.00%	90.00%
The missing rate	13.33%	13.33%	6.67%
False alarm rate	13.33%	6.67%	13.33%

The error curve in the network error training process of Group 3 is shown in Figure 4.

Ten characteristic parameters were used in time domain for SVM training and classification and the best classification results can be achieved at an accuracy rate of 90%. The missing rate and false alarm rate are below 15%, showing that the classification effect is better. The above classification results indicate that the bio-photon data features of the test group and control group in time domain contain a large amount of information about the wheat, proving that these features are the effective feature parameters of the classifier.
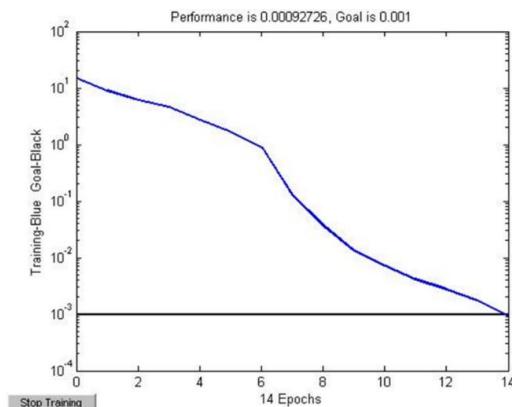
Figure 4. The error curve in the network error training process of group3

4. Conclusion

This paper integrates the time domain and frequency domain characteristics of the spontaneous emission signals of the wheat and uses the SVM classifier to do the preliminary classification. It is found that the classification accuracy of the SVM can be up to 93.33%. The results show that SVM is effective in the identification of these features and that the time domain features of the bio-photon signals of wheat contain a large amount of information and can be used to achieve effective classification. This provides certain experimental basis for the further study of the classifier model with the spontaneous emission information about the wheat with hidden pests.

References

Abdullahi N. K., Ahmad M. S., Abubakar A. (2018). Application of electrical resistivity tomography technique for delineation of gold mineralization in Bugai town, Birnin Gwari, Kaduna, North Western Nigeria. *Environmental and Earth Sciences Research Journal*, Vol. 5, No. 1, pp. 29-35. https://doi.org/10.18280/eesrj.050201

Alamdar F., Ghane S., Amiri A. (2016). On-line twin independent support vector machines. *Neurocomputing*, Vol. 186, pp. 8-21. https://doi.org/10.1016/j.neucom.2015.12.062

Bello A. A., Mamman M. B. (2018). Monthly rainfall prediction using artificial neural network: A case study of Kano, Nigeria. *Environmental and Earth Sciences Research Journal*, Vol. 5, No. 2, pp. 37-41. https://doi.org/10.18280/eesrj.050201

Chou J. S., Cheng M. Y., Wu Y. W. (2014). Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert System*, Vol. 41, No. 8, pp. 3955–3964. https://doi.org/10.1016/j.eswa.2013.12.035

Cortes C., Vapnik V. (1995). Support-vector networks. *Machine Learning*, Vol. 20, No. 2, pp. 273-297. https://doi.org/10.1023/A:1022627411411

Devos O., Downey G., Duponchel L. (2014). Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils. *Food Chemistry*, Vol. 148, pp. 124-130. https://doi.org/10.1016/j.foodchem.2013.10.020

Feng Y., Palomar D. P. (2015). Palomar Normalization of linear support vector machines. *IEEE Trans, Signal Process*, Vol. 63, No. 7, pp. 4673-4688. https://doi.org/10.1109/TSP.2015.2443730

Li G., You J., Liu X. (2015). Support Vector Machine (SVM) based prestack AVO inversion and its applications. *Journal of Applied Geophysics*, Vol. 120, pp. 60-68. https://doi.org/10.1016/j.jappgeo.2015.06.009

Liang Y., Song H., Liu Q., Shi W. Y., Li L. (2014). Study on spectrum estimation in biophoton emission signal analysis of wheat varieties. *Mathematical Problems in Engineering*, No. 2, pp. 1-9. https://doi.org/10.1155/2014/606275

Maldonado S. (2014). Feature selection for high-dimensional class-imbalanced data sets using support vector machines. *Information Sciences*, Vol. 286, pp. 228-246. https://doi.org/10.1016/j.ins.2014.07.015

Maldonado S., Pérez J., Weber R. (2014). Feature selection for support vector machines via mixed integer linear programming. *Information Sciences*, Vol. 279, pp. 163-17. https://doi.org/10.1016/j.ins.2014.03.110

Nalepa J., Siminski K., Kawulok M. (2015). Towards parameter-less support vector machines. *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, No. 3, pp. 211-215. https://doi.org/10.1109/ACPR.2015.7486496

Sun Y., Leng B., Guan W. (2015). A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing*, Vol. 166, pp. 109-121. https://doi.org/10.1016/j.neucom.2015.03.085

Thilina K. M., Choi K. W. N., Saquib N. (2012). Pattern classification techniques for cooperative spectrum sensing in cognitive radio networks. *SVM and W-KNN Approaches*, Vol. 12, No. 7, pp. 1260-1265. https://doi.org/10.1109/GLOCOM.2012.6503286

Van N. H., Patel V. M., Nasrabadi N. M. (2013). Design of non-linear kernel dictionaries for object recognition. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, Vol. 22, No. 12, pp. 5123-5135. https://doi.org/10.1109/TIP.2013.2282078

Wang F., Duan S., Liang Y., Shi W. (2014). Research on the ultra-weak luminescence of maize seeds. *Journal of Chemical & Pharmaceutical Research*, Vol. 6, No. 5, pp. 42-46.

Zhang K., Qian K., Chai Y. (2014). Research on fault diagnosis of Tennessee Eastman process based on KPCA and SVM. *International Symposium on Computational Intelligence & Design*, *IEEE*, Vol. 1, pp. 490-495. https://doi.org/10.1109/ISCID.2014.234