World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:6, No:7, 2012

Enhanced Performance for Support Vector Machines as Multiclass Classifiers in Steel Surface Defect Detection

Ehsan Amid, *Student Member, IEEE,* Sina Rezaei Aghdam, *Student Member, IEEE* and Hamidreza Amindavar, *Member, IEEE*

Abstract—Steel surface defect detection is essentially one of pattern recognition problems. Support Vector Machines (SVMs) are known as one of the most proper classifiers in this application. In this paper, we introduce a more accurate classification method by using SVMs as our final classifier of the inspection system. In this scheme, multiclass classification task is performed based on the "one-against-one" method and different kernels are utilized for each pair of the classes in multiclass classification of the different defects.

In the proposed system, a decision tree is employed in the first stage for two-class classification of the steel surfaces to "defect" and "non-defect", in order to decrease the time complexity. Based on the experimental results, generated from over one thousand images, the proposed multiclass classification scheme is more accurate than the conventional methods and the overall system yields a sufficient performance which can meet the requirements in steel manufacturing.

Keywords—Steel Surface Defect Detection, Support Vector Machines, Kernel Methods.

I. INTRODUCTION

AUTOMATIC surface inspection is widely used in steel manufacturing with the purpose of controlling the quality of the product. In this application, a line camera captures images of steel surfaces and a pattern recognition system is employed for classification of the captured images to their corresponding classes. Support Vector Machines (SVMs) are known as the most applicable classifiers in steel surface defect detection .Various successful implementation of SVMs have been reported in this application. SVM was developed in the study of Vapnik [1]. Jia [2] described a real-time visual inspection system, utilizing SVM for detection of defective patterns. The proposed system was found to be effective from the viewpoint of accuracy and speed. Choi [3] also adopted SVM as their final classifier of the inspection system and a good accuracy in detecting five defects on rolling strip was reported.

SVMs were originally single kernel classifiers which were designed for two class classification. Several methods have been proposed to extend SVMs for multiclass classification. Knerr [4] introduced "one-against-one" method and Hsu [5] compared this method with a couple of other methods and indicated that the "one-against-one" method is more suitable for practical use than the other methods.

H. Amindavar is with the Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran.

E. Amid and S. R. Aghdam e-mail: {ehsan_amid, sina.rezaei}@aut.ac.ir

SVMs employing different kernels yield different performances in detection of defects with different shapes. "scratch", "roll imprint", "edge strain" and "pit" are four kinds of common defects in steel manufacturing. In this study, we introduced a multiclass classification scheme based on one-against-one method, in which we employed various kernels in the classification of the mentioned defects. Aimed to do this, after training phase, four different kernels, including "linear", "radial basis functions", "quadratic" and "polynomial" were applied on the validation data, altogether and for each pair of the defects, the kernel which produced the highest accuracy was adopted as the final kernel for the classification of the corresponding defect in the test phase.

In order to decrease the time complexity of the classification task, a decision tree is used in the first stage for two-class classification of the images to "defect" and "non-defect". Then the proposed multiclass classification is performed on the defective images. Experimental results, proved that the proposed multiclass classification task is more accurate than the conventional methods and the overall classification system is absolutely suitable in practice, for meeting accuracy and speed requirements.

The proposed approach is presented in the rest of paper, as follows: In section II, the basics of the SVM as a popular classifier in steel surface defect detection is reviewed. In section III, four typical examples of kernels used in pattern recognition applications are presented. In section IV, decision trees are described as a class of rapid classifiers. The feature extraction scheme utilized in this experiment which is based on Local Binary Pattern is presented in section V. Section VI discusses the experimental results for the proposed approach. And section VII gives the conclusion.

II. SUPPORT VECTOR MACHINES

SVM is a learning machine originally utilized for two-class classification problems and has emerged as one of the most popular classifiers in various applications. In this classification scheme, input vectors are firstly mapped into a high-dimensional space through a nonlinear mapping and then a hyperplane is constructed and is moved until an appropriate separation is achieved by the hyperplane that leaves the maximum possible margin from both classes.

So the goal is to construct such a hyperplane from the

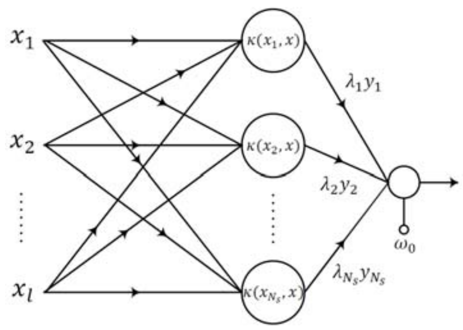


Fig. 1. The SVM Architecture Employing Kernel Functions. [6]

training samples. Given that $T = \{(x_1, y_1), , (x_n, y_n)\}$ is the training set, $\mathbf{x}_i \in R^n$ are input vectors and $y \in \{-1, 1\}$ are classification labels, the desired hyperplane is

$$\boldsymbol{\omega}^T \mathbf{x} + b = 0 \tag{1}$$

SVMs require the solution of the following optimization problem

$$\underbrace{\min}_{\boldsymbol{\omega}, b} \frac{1}{2}||\boldsymbol{\omega}||^2 + C\sum_{i=1}^{n} \xi_i$$

$$\text{S.T} \quad y_i(\boldsymbol{\omega}^T \mathbf{x}_i + b) \geq 1 - \xi_i \tag{2}$$

While we stated the original problem in a finite dimensional space, it often happens that in such a space, the datasets to be classified are linearly inseparable. Hence it was offered that the original finite dimensional space be mapped into a much higher dimensional space, where the classes can satisfactorily be separated by a hyperplane.
Kernel functions $\kappa(x, z) = \sum_r \phi_r(x)\phi_r(z)$ are defined to suit the problem. Typical examples of kernels used in pattern recognition applications are:

1) Linear kernel function
2) Quadratic kernel function
3) Polynomial kernel function
4) Radial Basis Function (RBF)

Once the kernel function $\kappa(x, z)$ is adopted the optimization task, (2) now becomes

$$\underbrace{\max}_{\lambda} \left(\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right) \tag{3}$$

$$\text{S.T} \quad 0 \leq \lambda_i \leq C, \quad i = 1, 2, \ldots, N \tag{4}$$

$$\textstyle\sum_i \lambda_i y_i = 0 \tag{5}$$

Where, $\lambda_i \geq 0$ are the Lagrange multipliers. After solving the optimization problem, the resulting linear classifier is obtained:

$$g(x) = sgn\left(\sum_{i=1}^{n} \lambda_i y_i \kappa(x_i, x) + \omega_i \right) \tag{6}$$

Fig. 1 shows the corresponding architecture. ([6],[7])

III. Kernel Types

Different kernel types which are used in this paper are as follows:

A. Linear Kernel

The linear kernel function $\kappa_l(\mathbf{x}, \mathbf{z})$ is simply dot product of the vectors $\mathbf{x}$ and $\mathbf{z}$ in Euclidean space

$$\kappa_l(\mathbf{x}, \mathbf{z}) = \mathbf{x} \cdot \mathbf{z} \tag{7}$$

Linear kernel is used in case of linearly separable classes where a hyperplane can be found with maximum margin from the nearest point of each class. It does not apply any mapping over the feature space and therefore, does not include any parameters to be adjusted.

B. Quadratic Kernel

Quadratic kernel function is also a non-parametric method which is given by

$$\kappa_q(\mathbf{x}, \mathbf{z}) = \mathbf{x} \cdot \mathbf{z}(1 + \mathbf{x} \cdot \mathbf{z}) \tag{8}$$

It is a very popular kernel method in different applications due to its simplicity and efficiency.

C. Polynomial Kernel

The polynomial kernel over a vector space X of dimension n is defined as

$$\kappa_d(\mathbf{x}, \mathbf{z}) = (\langle \mathbf{x}, \mathbf{z} \rangle + R)^d \tag{9}$$

Where R and d are kernel parameters. It can be shown that the dimension of the feature space for the polynomial kernel is

$$\begin{pmatrix} n+d \\ d \end{pmatrix} \tag{10}$$

Moreover, using the binomial theorem we can expand (7):

$$\kappa_i(\mathbf{x}, \mathbf{z}) = \sum_{s=0}^{d} R^{d-s} \langle \mathbf{x}, \mathbf{z} \rangle^s \tag{11}$$

It is apparent that by decreasing R, the relative weighting of the higher order polynomial increases. A proof for equation (10) is available in [8].

D. Gaussian Kernel

The Gaussian kernel is defined by

$$\kappa(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{||\mathbf{x} - \mathbf{z}||^2}{2\sigma^2} \right) \tag{12}$$

It is also referred as radial basis function as it only depends on the distance of $\mathbf{x}$ and $\mathbf{z}$ in the input space. The parameter σ acts like degree d in the polynomial kernel controlling the flexibility of the kernel where small values of σ corresponds to the large values of d. If σ is chosen too small, it increases the risk of overfitting while large values of σ reduces the kernel to a constant function which is not applicable to learn any significant classifiers.

TABLE I
THE BASIC ALGORITHMIC STEPS FOR A TREE. [6]

Begin with the root node, i.e., $X_t = X$
For each new node t
  For every feature $x_k, k = 1, 2, \dots, l$
    For every value $\alpha_{kn}, n = 1, 2, \dots, N_{tk}$
      • Generate X_{tY} and X_{tN} according to the answer in the question: is $x_k(i) \leq \alpha_{kn}?, i = 1, 2, \dots, N_t$
      • Compute the impurity decrease
    End
    • Choose α_{kn_0} leading to the maximum decrease w.r.t. to x_k
  End
  • Choose x_{k_0} and associated $\alpha_{k_0 n_0}$ leading to the overall maximum decrease of impurity
  If the maximum value of $\Delta I(t)$ over all possible splits is less than a threshold value
  $\rightarrow$ stop splitting and assign a class
  If not
  $\rightarrow$ generate two descendant nodes t_Y and t_N with associated subsets X_{tY} and X_{tN}, depending on the answer to the question: is $x_{k_0} \leq \alpha_{k_0 n_0}$?
End

IV. DECISION TREES

Decision trees are a class of nonlinear classifiers in which classes are sequentially rejected until a finally accepted class is attained. The sequence of decisions is applied to individual features, and the questions to be answered are of the form ”is feature $x_i \leq \alpha$?”, where α is a threshold value.
For each node, t, splitting is equivalent with the division of the subset X_t into two disjoint descendant subsets, X_{tY} and X_{tN} corresponding to the answers ”Yes” and ”No” of the questions respectively.
In order to adopt a splitting rule, we define a parameter called node impurity, denoted as $I(t)$, given by

$$I(t) = -\sum_{i=1}^{M} P(\omega_i|t) \log_2 P(\omega_i|t) \tag{13}$$

And decrease in node impurity is defined as

$$\Delta I(t) = I(t) - \frac{N_{tY}}{N_t} I(t_Y) - \frac{N_{tN}}{N_t} I(t_N) \tag{14}$$

where N_{tY} and N_{tN} are the number of points being sent to the subset X_{tY} and X_{tN}.
The general recursive procedure for the construction of decision tree is shown in Table I, in which a commonly used rule for class assignment is the majority rule, i.e., the leaf is labeled as ω_j where

$$j = arg \underbrace{max}_{i} P(\omega_i|t) \tag{15}$$

V. FEATURE EXTRACTION

There have been significant advances in feature extraction methods for surface defect detection in the recent years. In this paper, we have focused on a feature extraction scheme based on texture analysis method.

p=1	p=2	p=3
p=8	c	p=4
p=7	p=6	p=5

Fig. 2. Derivation of the Original LBP.

r=1			p=1			s=6
	r=2		p=2		s=5	
		r=3	p=3	s=4		
q=1	q=2	q=3	c	q=4	q=5	q=6
		s=3	p=4	r=4		
	s=2		p=5		r=5	
s=1			p=6			r=6

Fig. 3. Derivation of the NLBP.

A. Local Binary Pattern

Local Binary Pattern (LBP) is a general definition of texture in a local neighbourhood which is introduced by Ojala. [7]. Two dimensional distribution of LBP can be used as features due to their good discrimination rates and LBP operator is known as a simple rotation invariant method for feature extraction.
The derivation of LBP is represented in [8]. The mathematical definition of the original LBP, according to Fig. 2, is as follows:

$$LBP_{P,R} = \sum_{p=0}^{P-1} sgn(g_p - g_c) 2^p \tag{16}$$

where g_c and g_p are the gray level of the center pixel and neighbour pixels, respectively and *sgn(.)* is the sign function, defined as

$$sgn(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{17}$$

B. Feature Extraction Approach based on New LBP

In order to introduce an extended LBP based operator, we consider a 7×7 window, as shown in the Fig. 3, and define U as a uniformity criterion which is described as the number of alterations from ”1” to ”0” and vice versa in the sequence. For example for the sequences 00000000 or 11111111, uniformity criterion (U) is equal to zero and for the binary form of the numbers 2, 4, 8 and 16, it is equal to 2, since there are two 0/1 alterations in the mentioned sequences. The new LBP is, then, derived from:

$$NLBP = \begin{cases} \sum_{p=0}^{P-1} sgn(g_p - g_c) & \text{if } U(\mathit{NLBP}) \leq K \\ P+1 & \text{Otherwise} \end{cases} \tag{18}$$

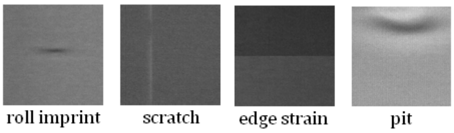


Fig. 4. Four Common Defect Types in Steel Industry.

TABLE II
DIVISION OF THE IMAGE DATASET INTO TRAIN, VALIDATION AND TEST SETS.

Defect	Train	Validation	Test	Total
Scratch	50	50	70	170
Roll Imprint	50	50	70	170
Edge Strain	50	50	70	170
Pit	50	50	70	170
Percentage	50 (%29.41)	50 (%29.41)	70 (%41.18)	680

TABLE III
RESULTS FOR MULTICLASS CLASSIFICATION OF THE DEFECTS.

Defect	Accuracy
Scratch	%92.94
Roll Imprint	%96.47
Edge Strain	%95.88
Pit	%97.64
Total	%95.73

where

$$U(NLBP) = \sum_{g=p,q,r,s} \sum_{t=1}^{6} |sgn(g_t - g_c) - sgn(g_{t-1} - g_c)| \quad (19)$$

VI. EXPERIMENTAL RESULTS

Our database consists of 1080 100 $\times$ 100 pixel images achieved from the production line of the "Mobarake Steel Manufacturing" and preprocessed by "Dideh Pardaz Saba Company". The database contributes 680 defect images which belong to four common defect types in steel industry: scratch, roll imprint, edge strain, and pit. These defects are shown in Fig. 4. These defect images were divided into training, validation and test sets, containing 30, 30 and 40 percent of the total, respectively. Then, we extracted a feature vector containing 196 features by applying NLBP algorithm introduced in section V with K=10 to each image. The training set for each defect type was used to train SVMs with different kernel type based on one-against-one method. In this experiment, we used four popular kernel types: linear, polynomial, Gaussian, and quadratic. This brings about four different SVMs for each two-class classification. Then, by applying the validation set, the SVM with maximum accuracy for that two-class classification was selected from the group of four SVMs for each combination of two defects. This resulted in SVMs with different kernel types for the classification purpose between different defects. Finally, the system was applied to the test set to evaluate the overall performance. The whole process of the defect detection based on multi-kernel SVMs is presented in Fig. 5. The results are shown in Table III.

To compare the accuracy of our proposed method with the simple classification method by using the same kernel type

TABLE IV
MULTICLASS CLASSIFICATION OF DEFECTS.

Classification Method	Accuracy	Time
Traditional Multiclass SVM	%91.61	2.9129
Proposed Method	%95.73	4.1215

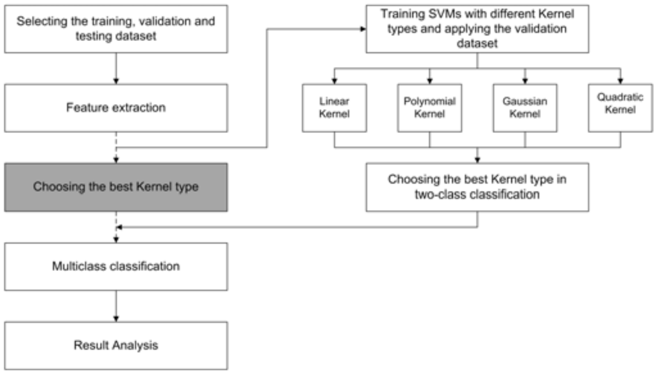


Fig. 5. The Whole Process of Defect Classification.

for each pair of classes, we applied the training and test sets to a system with SVMs based on only linear, polynomial, RBF or quadratic kernel type. The results are shown in Table IV. The results indicate that our proposed method yields better performance in classification of different defect types.

A drawback associated with the proposed method is its relatively higher time complexity. Aimed to tackle this problem, we proposed a two-stage classification system, shown in Fig. 6, in which we utilized decision tree as a rapid classifier in the first stage, for two-class classification of the images to "defect" and "non-defect". Since the multiclass classification task is merely applied on the surfaces which are recognized as defective by the two-class classifier of the first stage, a significant decrease in executing time is achieved. Table V represents the accuracy and executing time of the proposed classification system when employed on 1080 images including 400 non-defect and 680 defective images. The results validated that the presented scheme yields a better performance than the conventional multiclass SVM classifier. The time stated in the table is the total time in seconds, required for the classification of 1080 100$\times$100 pixel images on a 2.4 GHz PC.

VII. CONCLUSION

Traditional SVMs employing "one-against-one" scheme for multiclass classification problems, use to apply a single predefined kernel. The mentioned kernel, however, may not yield the best possible performance in classification of different defects.

In this paper, we introduced a classification method, in which for each pair of different defects, a kernel which produced the highest accuracy in the validation phase, was adopted as the final kernel for the classification of the corresponding pair of defects in the test phase.

To testify the effectiveness of the proposed method, simulations are conducted on features being extracted using

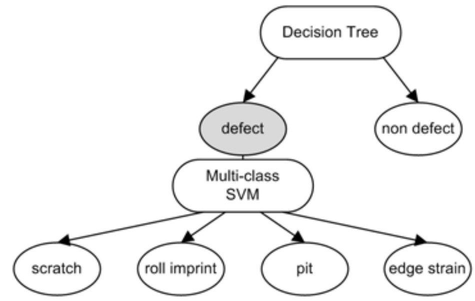


Fig. 6. Proposed Classification Architecture.

TABLE V
MULTICLASS CLASSIFICATION OF COMPLETE IMAGE DATASET INCLUDING DEFECT AND NON-DEFECT.

Classification Method	Accuracy	Time
Traditional Multiclass SVM	%90.37	4.6217
Proposed Two-stage Classification Scheme	%97.68	5.1269

an extended LBP based operator applied on more than a thousand of steel strips. The simulation results indicated that the proposed method can accomplish a higher accuracy in multiclass classification of the steel surface defects, rather than the conventional method, in favor of the increased executing time.

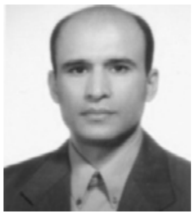
In order to tackle the problem of increased time complexity, a two stage classification architecture is proposed, in which a decision tree is utilized in the first stage, for two-class classification of the images to "defect" and "non-defect". The simulation results proved that the proposed classification system can yield a better accuracy rather than the traditional methods, in a sufficient time which can meet the speed requirements in steel manufacturing.

ACKNOWLEDGMENT

The authors would like to thank Mohammad Faghih Imani, Navid Rabbani and the Dideh Pardaze Saba Co. for providing the preprocessed data. The authors would also like to thank Hamed Delavash Gargari for his assistance.

REFERENCES

[1] V. Vapnik, *Statistical Learning Theory*, Wiley, 1998.
[2] H. Jia, Y. L. Murphey, J. Shi, T. Chang, "An intelligent real-time vision system for surface defect detection", *Proceedings of the 17th International Conference on Pattern Recognition*, 2004. 239 - 242
[3] K. Choi, K. Koo, J. S. Lee, "Development of defect classification algorithm for POSCO rolling strip surface inspection system", *International Joint Conference SICE-ICASE*, 2006. 2499 - 2502
[4] S. Knerr, L. Personnaz, G. Dreyfuss, "Single-layer learning revisited: a stepwise procedure for building and training a neural network", *In Neurocomputing: Algorithm, Architectures and Applications (1990)*, p. 41 - 50.
[5] C. Hsu, C. Lin, "A Comparison of methods for multiclass support vector machines", *IEEE Transactions on Neural Networks*, 2002. 415 - 425
[6] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. 3rd Edition, Academic Press, 2006.
[7] C. Shi, Y. Wang, H. Zhang, "Faults diagnosis based on support vector machines and particle swarm optimization", *International Journal of Advancements in Computing Technology*, Vol. 3, No. 5, pp. 70 $\sim$ 79, 2011.
[8] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.



Hamidreza Amindavar received the B.Sc., M.Sc., and Ph.D. degrees from the University of Washington in Seattle, in 1985, 1987, and 1991, respectively, all in electrical engineering. He also holds an M.Sc. degree in applied mathematics from the same university.
He is currently an Associate Professor in the Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran. His research interests include signal and image processing, array processing, and multiuser detection.
Dr. Amindavar is a member of Tau Beta Pi and Eta Kappa Nu.

Ehsan Amid received the B.Sc. degree in electrical engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2012.
His research interests include speech processing, image processing and pattern recognition.

Sina Rezaei Aghdam received B.Sc. degree in electrical engineering from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2011. He is currently working toward the M.Sc. degree at the Amirkabir University of Technology (Tehran Polytechnic).
His research interests range from wireless communications to digital signal processing.