Int. j. inf. tecnol.
https://doi.org/10.1007/s41870-017-0080-1

ORIGINAL RESEARCH

Survey on SVM and their application in image classification

Mayank Arya Chandra[1] · **S. S. Bedi**[1]

Received: 11 May 2017 / Accepted: 22 December 2017

Abstract Life of any living being is impossible if it does not have the ability to differentiate between various things, objects, smell, taste, colors, etc. Human being is a good ability to classify the object easily such as different human face, images. This is time of the machine so we want that machine can do all the work like as a human, this is part of machine learning. Here this paper discusses the some important technique for the image classification. What are the techniques through which a machine can learn for the image classification task as well as perform the classification task with efficiently. The most known technique to learn a machine is SVM. Support Vector machine (SVM) has evolved as an efficient paradigm for classification. SVM has a strongest mathematical model for classification and regression. This powerful mathematical foundation gives a new direction for further research in the vast field of classification and regression. Over the past few decades, various improvements to SVM has appeared, such as twin SVM, Lagrangian SVM, Quantum Support vector machine, least square support vector machine, etc., which will be further discussed in the paper, led to the creation of a new approach for better classification accuracy. For improving the accuracy as well as performance of SVM, we must aware of how a kernel function should be selected and what are the different approaches for parameter selection. This paper reviews the different computational model of SVM and key process for the SVM system development. Furthermore provides survey on their applications for image classification.

Keywords Support vector machine (SVM) · Image classification · Kernel function · Machine learning · Color image · Decision space

✉ Mayank Arya Chandra
machandra100@gmail.com

S. S. Bedi
erbedi@yahoo.com

[1] Department of CSIT, MJP Rohilkhand University, Bareilly, UP, India

1 Introduction

A million kinds of living things are in this world. Then, how can anyone distinguish among them. So it is important to all living being must be differentiate among various objects, color, smell to every day of life. For any activity to be performed by any living being, expertise must be possessed to differentiate between various inputs they receive from the external world through their sensory organs. Every day to day task starting from morning, human being differentiates the object from another object easily, but if there is a machine then it is as easy as human, obviously not. For making a machine as capable as human in the field of object classification, there are lots of technique has been developed, which gives the best performance in differentiating the object. There is a lot of a classification technique for data classification. However, no single classification technique has been found to best accuracy for all kinds of data set. SVM is increasing attention because it has pure mathematical foundation and seems to accomplish to a certain extent excellently in various diverse real-world applications. Inspiration of their increased classification accuracy in data mining and image classification, with SVM, here, this paper discusses the different approaches of image classification with SVM.

SVM is a powerful mathematical computational model for classification task. SVM is a supervised learning approach which is applied in the field of regression and classification [1]. It has solid statistical foundation and effectiveness. Over the past decades, various alterations to SVM appeared, such as Twin SVM [2], Lagrangian SVM [3], least square SVM [2], Decision tree SVM, DAG SVM, multi kernel classification [4]. SVM has two most popular approaches for translating Multi-class classification into multiple binary classifications is OvO (One vs One) and OvR (One vs Rest) [5]. For N-class classification problem OvO require (N × (N-1))/2 binary classifiers. OvR has single classifiers per class with dataset of that class as positive dataset and all other datasets as negatives. For such comparison OvR implement a series of binary classifiers where each classifier separates one class from the rest of the class [6]. OvR requires N binary comparison for each test data. As compared to OvO, OvR has less computational complexity.

Today scenario SVM used with other computational model, this increase the accuracy and reduce complexity. It is well known that not any single classification technique has been implemented to be superior over all datasets. For increasing the classification accuracy different article uses different technique some them hybrid approach which give the better performance over a single technique. This paper discusses the different approaches used with SVM for image classification. This paper organizes the different section. First section gives the general view of SVM in the introductory part. Section 2 gives briefly about the SVM and its supporting techniques. Section 3 deals the recent technique for image classification, and then concludes the paper at the end of the section.

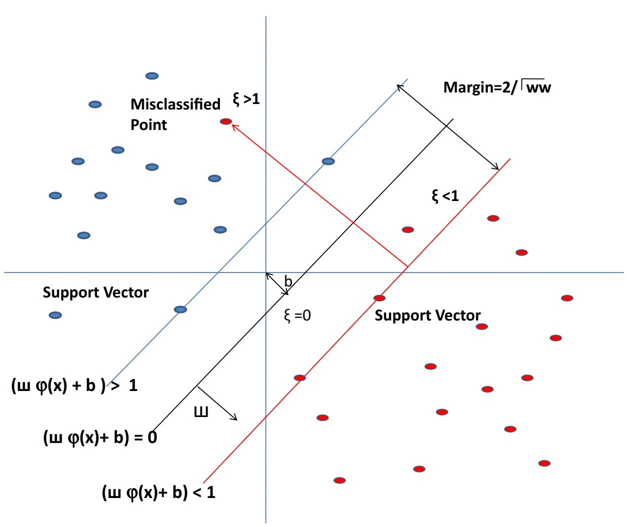

Fig. 1 SVM separable problem in 2-dimensional space [3]

2 SVM and supporting techniques

1990′s SVM theory had been developed by Vapnik and his group. SVM concept was inherited from neural network or may be saying that SVM is mathematical extension of Neural Network. SVM can classify both linear and non-linear data [7]. SVM performs classification by transforming the original training data into multidimensional space and constructing hyper-plane in higher dimensional (Fig. 1). SVM is an optimal hyper-plane based mathematical learning scheme [1].

An SVM performs classification by constructing a hyper-plane in higher dimensions. This hyper-plane refer as decision planes shown in Fig. 2. A decision plane is one that distinguishes a group of data of one type from another type. A suitable nonlinear mapping for higher dimensional space always classifies a data by constructing a hyper-plane [8]. SVM search those vector points, refer as support vector, which define the decision boundary and give the large marginal separation between the classes [9]. SVM separates the classes with maximum marginal distance in the decision plane [9–11].

Middle line represents the maximum-margin hyperplane. In other words, one would select boundary line that separates the two classes at a maximal distance to the closest data point [12]. For two class problem, decision boundary or hyper-plane for classification is defined by Vapnik's theory. If input set is x_i is $m * n$ dimension and corresponding class is $y_i \in \{-1, +1\}$.

$$w^T \emptyset(x_i) + b = 0 \tag{1}$$

The hyperplane is built to such that it fulfills the following inequality function for both the classes.

$$w^T \emptyset(x_i) + b \geq +1 \qquad y_i = +1 \tag{2}$$

$$w^T \emptyset(x_i) + b < +1 \qquad y_i = -1 \tag{3}$$

By using Eq. (1) and Eq. (2) we get Eq. (4).

$$y_i(w^T \emptyset(x_i) + b) \geq +1 \qquad i = 1.\ldots.N \tag{4}$$

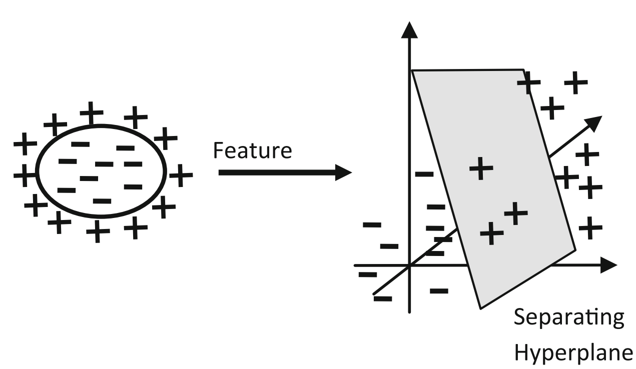

Fig. 2 Decision space [10]

Margin is the smallest perpendicular distance to the data point from the hyper-plane. The Maximum marginal hyper-plane is the decision plane where the margin is largest. SVM selects the maximum margin separating hyper-plane. Maximum marginal hyper-plane selection done by SVM, increase the ability of accurate classification and reduces the chances of misclassification [13]. Two hyper-planes D1 and D2 are shown in the Fig. 3. Figure shows that hyper-plane D1 is better than hyper-plane D2 because it maximizes the margin. For analyzing the Fig. 3 the accepting hyper-plane is D1, if any, data points that lying on the wrong side or within the margin of correct side of the hyper-plane. To resolve such type of cases, the SVM algorithm has introduced the new term known as 'soft margin' [14]. Data has noise and outliers this problem is solved in different ways by different flavors of SVM; here slack variable concept is used:

Without "slack variable":

$$y_i(w^T x_i + b) \geq 1 \tag{5}$$

With "slack variables": $S_i \geq 0$ and allow

$$y_i\left(w^T x_i + b\right) \geq 1 - S_i \tag{6}$$

Now considering the convex optimization problem

$$\min_{w,b,S} f(w,b,f) = \frac{1}{2} w^T w + c \sum_i S_i. \tag{7}$$

Lagrange gives the new terms known as saddle point which gives the solution of optimization problem represented in Eq. (8)

$$y(x) = \text{sign}\left[\sum_{i-1}^{n} \alpha_i y_i K(x_i x) + b\right] \tag{8}$$

Where $K(x_i x) = \emptyset(x_i)^T \emptyset(x)$ is a kernel function that must satisfy the Mercer Theorem.

One of the best concepts used in SVM is kernel function. The kernel function is a solid mathematical method used for nonlinear mapping for higher dimensional data [11]. Through this SVM can solve the higher dimensional

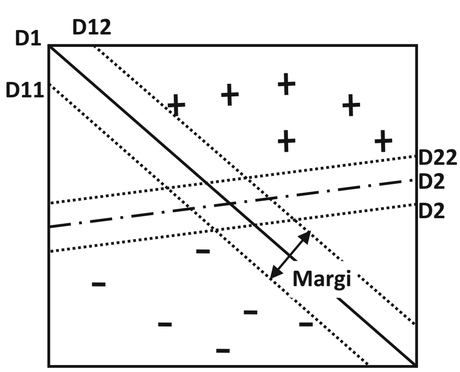

Fig. 3 Maximum margin hyper-plane [10]

classification of set of originally input data space [15]. In general, it calculates the value of the dot product mapped data point into feature space. The benefit of kernel function is that the complexity of the problem only dependent on the dimensionality of input space rather than feature space. Some of popular kernel functions are.

1. Linear Kernel Function: Simple kernel function known as Linear kernel function.

$$K\left(x, x'\right) = x \cdot x^T \tag{9}$$

Where x^T is a transpose of the input matrix x 'Lin' keyword used in MATLAB for implementing the linear function. If we have too many feature, then over fitting can happen, this is the drawback of this Linear Kernel Function.

2. Polynomial Kernel Function: It is a non-linear kernel function. It is directional that means output depends on the direction of input in a low dimensional space. This is due to the dot product in the kernel [16].

$$K(x, x_i) = \left(1 + x \cdot x_i^T\right)^P \tag{10}$$

Where, 'p' is the degree of kernel function.'Poly' keyword used in MATLAB for implementing the Polynomial Function.

3. Radial Basis Kernel Function: Radial basis kernel function is widely used with SVM [15]. Its select the solution those are smooth.

$$K(x, x_i) = e^{-\gamma x - x_i^2} \gamma > 0 \tag{11}$$

Where, γ is a parameter that sets the spread of the kernel.'rbf' keyword used in MATLAB for implementing the Radial Basis Kernel Function.

3. Sigmoid Function: The hyperbolic tangent function known as a Sigmoid Function.

$$K(x, y) = \tanh\left(ax^T y + c\right), a, c > 0 \tag{12}$$

Here 'a' and 'c' is kernel parameters.Due to limitation of every kernel function for different data set, it is an existing research area to develop a novel kernel function that will provide the more efficiency and accuracy. The difficulty of designing a universal kernel function which actually an emerging research topic in machine learning [17].

2.1 Wavelet

Wavelet is a powerful mathematical function used to analyze data and extract information from different data set. The term wavelet comes from the fact that they integrate to zero; they wave up and down across the axis. In 1909 Wavelet first referenced by Alfred Haar on his Ph D Thesis titled "On the theory of the orthogonal function systems" [18]. In the early 1980s, David Marr use wavelet concept in artificial vision for robots [19].M A Chandra uses wavelet for hiding data in video [20]. Victor Wickerhauser uses the wavelet for compression of sound [21]. Wavelet theory has been applied to several subjects. Wavelet transform are now being adopted by a number of areas of signal processing, DNA analysis, protein analysis, image processing, data hiding, and so on.

In recent years, wavelet theory developed rapidly in every field of computer science. Mother Wavelet can be formed as eq. (13).

$$\psi_{a,b}(x) = |a|^{1/2}\psi\left(\frac{x-b}{a}\right) \tag{13}$$

Where, ‘a’ is the dilatation factor and b is the translation factor. Simplest wavelet is haar wavelet, which is widely used in signal, image analysis. It’s dilation and translations of the function is

$$\psi_{a,b}(x) = 2^{a/2}\psi(2^a x - b) \tag{14}$$

The haar function is pairwise orthogonal in $L^2(R)$. The haar2 $\times$ 2 matrix is given by

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{15}$$

Due to the orthonormal property, the wavelet coefficients of a signal f(x) can be easily computed via Eq. (16)

$$c_{a,b} = \int_{-\infty}^{+\infty} f(x)\psi_{a,b}(x)dx \tag{16}$$

And the synthesis formula

$$f(x) = \sum_{a,b} C_{a,b}\psi_{a,b}(x) \tag{17}$$

Equation no (17) can be used to recover $f(x)$from its wavelet coefficients. To construct the mother wavelet $\psi(x)$ we may first determine a scaling function(x).

$$\varphi(x) = \sqrt{2}\sum_k h(k)\varphi(2x-k) \tag{18}$$

Then, the wavelet kernel $\psi(x)$ is related to scaling function via

$$\psi(x) = \sqrt{2}\sum_k g(k)\varphi(2x-k) \tag{19}$$

Where

$$g(k) = (-1)^k h(1-k) \tag{20}$$

Wavelet transform generate the new possibility to extract image features at different scales. Discrete Wavelet Transform (DWT) is another strategy which is utilized to extricate the wavelet coefficients from various mother wavelet functions. DWT decomposed images into four subbands as shown in Fig. 4a, b. These sub bands labeled LH1, HL1 and HH1 represent the finest scale wavelet coefficients i.e., detail images while the sub-band LL1 corresponds to coarse level coefficients i.e., approximation image. Figure 4b represents the Three Scale DWT.

Wavelet is broadly used in analysis of image components, feature extraction, image segmentation, image compression, removing the noise, etc. Wavelet transformation is not well explored in image classification but extract image features at different scales which have more discriminative capability than the original spectral feature [22]. Wavelet has a powerful mathematical model for extracting discriminative features for high dimensional data set. A discriminative feature of the data is expressed by few wavelet coefficients [23, 24]. WT has been applied to a number of problems regarding data analysis and feature extraction [21, 25].

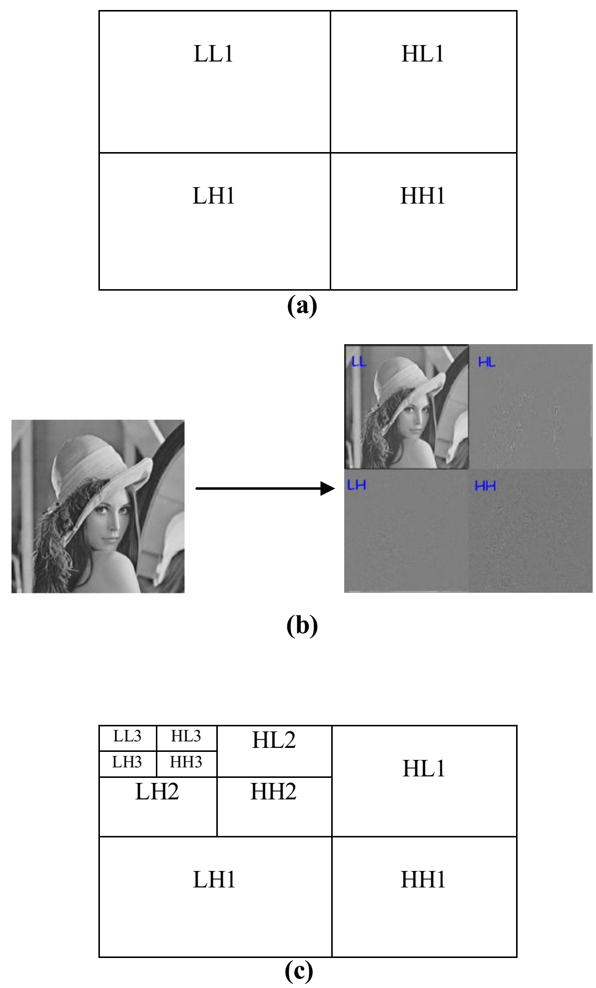

Fig. 4 **a** Image Decomposition. **b** Original lena image, lena image after wavelet decomposition. **c** Image decomposition different scale DWT

2.2 Neural network

Neural Network is an important tool for classification. it is machine learning model manifestation of human neural network. It is introduced by McCulloch and Pitts [26]. Neuron is the basic unit of neural network and carries information from one end to another.

A neural network consists of interconnected artificial neurons. Each neuron in a neural network has connection strength as associated weights $w_1, w_2, w_3 \ldots w_m$ the activation of neuron depends on input values and the corresponding weight. Weighted sum is calculated by summation function and their sum is passed through an activation function. Neurons are arranged in layers represented in Fig. 5.

The role of Activation functions is to limit the amplitude of the neuron output. When a neuron fires, the signal propagates from starting node to end node and calculate the output values.

$$v = \left(\sum_{j=1}^{m} w_j x_j + b \right) \tag{21}$$

Where v is induced field. Output of the function $y(m)$ is shown in Eq. (22)

$$y(m) = \varphi \left(\sum_{j=0}^{m} w_j x_j + b \right) \tag{22}$$

Where $\varphi()$ is activation function. This equation can be rewrite using sigmoidal activation function as in Eq. (23).

$$y(m) = sgn\left[w^T(m)x(m) + b \right] \tag{23}$$

Mid 50′s frank Rosenblatt created the perceptron concept based on supervised learning approach. Perceptron used as binary classifiers. It is a simple neuron that is used to classify its input into one of two classes [27]. Following Eq. (24) represent the two class classification

$$y = \begin{Bmatrix} +1 & w^T x + b \geq 0 \\ -1 & w^T x + b < 0 \end{Bmatrix} \tag{24}$$

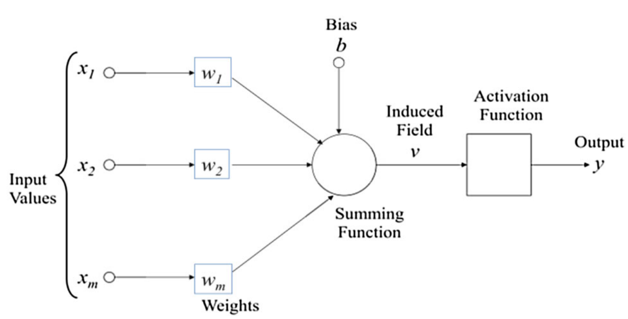

Fig. 5 Artificial Neuron Model

For minimizing the classification error, modify the weights $\boldsymbol{w}$ by using the update rule

$$w(m+1) = w(m) + \eta[d(m) - y(m)]x(m) \tag{25}$$

Where η learning parameter
d(n) Desired Output
y(n) Actual Output

Perceptron has limitation in multiclass classification, which is further resolved by werbos. Werbos introduced a concept of back-propagation. Back-Propagation algorithm is an extension of perceptron model.it has the same capability as the perceptron model.it optimizes the synaptic weight of neuron based on minimization of total error of the network over the set of training example. In 80′s Kohonen use the new term SOM (Self Organization Map). It is unsupervised learning concept use in visualizing the high dimensional data in low dimensional views [28]. In 1998 Broomhead and Lowe develop the RBF network based on radial basis function as an activation function [29]. From this it can be stated that ANN is a powerful computational model for solving real word problem in all applicable fields in engineering.

The neural network can produce accurate results. The major disadvantage in using ANN is to find the most appropriate grouping of training, learning and transfer function for classifying the data sets with growing number of features and classified sets. However, one of the main drawbacks is that it has like a "black-box" approach. They do not provide any insight into the underlying nature of the phenomena.

3 Recent technique

A huge amount of classification work could be found in literature. This section gives an overview of new SVM techniques in the field of the image class. Different SVM approaches and model named here, their precise description is presented in this section as well. Here a Table 1 describes the comprehensive summary of their approaches and their field. The classification problem main issue is how to classify the image data, how to categorize the image data, how to increase efficiency and accuracy of classification. Many researchers have been done for this, here we summaries the current research work in this domain.

3.1 Early 2000's

Early 50's, R.A. Fisher gives the first concept for the pattern recognition. In 58's, Rosenblatt introduces a concept of perceptron for linear classification [1]. In 60's Widrow and Hoff present the new concept known as Least-Mean-

Table 1 Approaches to using multiple features of image data for improving classification accuracy

Field	Feature/approaches	References
Hyperspectral data	'Spatially Uncorrelated (SU) test dataset and Spatial Correlated (SC), 242 bands covering the 400–2500 nm portion of the spectrum, Un-calibrated and noisy bands that cover water absorption features were removed, and the remaining 145 bands were included as candidate features	Chi et al. [36]
	Scattering feature, spectral-spatial information, multiscale and multidirection co-occurrence information	Yuan.et al. [48]
	Spectral and texture features, spectral-spatial information, smaller training sets	Kumar et al. [49]
	Region kernel. region-to-region distance similarity, point-to-point similarity	Peng et al. [51]
	Limited labeled data, high-level features, extract preliminary features, e.g., color and textures. Semi-supervised ensemble projection	Yang et al. [52]
	NVidia GPU platforms, composite kernel implementation, CUDA programming	Wu. et al. [53].
Chemical Pattern Data	Fast pruning algorithm, linear correlation analysis, principal component analysis (PCA), classification correlative component analysis (CCCA) of kernel matrix	Tao et al. [37].
	Apple core edge, Mould core, Un-mould core, Laplacian of Gaussian	Yang et al. [41]
Medical Image MRI Data	Structure of brain tissues, gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), vector field convolution (VFC) active contour	Tanoori et al. [38]
	Grain features, Gray level co-occurrence matrix (GLCM)	Liu et al. [40]
	Breast tissues,mean, SD	Lo. et al. [44]
	Colon tissue images, Laws' texture energy, Neighborhood gray-tone difference matrix, WEKA tool	Oliva et al. [59]
	Content based activecontour (CBAC) model, intensityand texture feature, Genetic Algorithm for optimizing the feature	Sachdeva et al. [62]
	First and second, derivative features	ImèneGarali et al. [63]
Fault diagnosis	Histogram features, Statistical features e.g., mean, standard error, median, standard deviation, sample variance, kurtosis, skewness, range	Sugumaran et al. [39]
Image classification	Multiview matrix completion, labeled set, average precision (AP) loss	Luo et al. [55]
	Single-layer sparse autoencoder, pooling operation, parameters are optimized by PSO and PCA	Yin et al. [56]
	Low level features (color, shape and texture, color-correlogram, wavelet transformation, Gabor wavelet	Sugamya et al. [57]
	Face feature parameters, Nose, Mouth, eye, eyebrow	Liu et al. [58]
	Pooling local information, average pooling, max-pooling, dense sampling, MAX, mean, region Ranking, region-score vector	Wei et al. [60]
	Low-level feature extracting HOG features, local object appearance and shape, distance metric	Ding et al. [61]
	Interactive image segmentation, geodesic distance features	Park et al. [64]
Plant leaves	Digital morphological features, PCA, support vector machine	Priya et al. [67]
	Color, shape and texture features, Two SVM classifiers	Es-saady et al. [68]
	Direct acyclic graph based multi-class least squares twin support vectormachine, hybrid feature selection (HFS) approach	Divya et al. [69]
	Support vector machine, hu moments and local binary patterns, histogram parameters	Marko Lukic [70]

Square (LMS) algorithm [30]. Werbos describe the back propagation method with gradient evaluation in 1974. 86's Back propagation term was popularized by Rumelhart and McClelland [31]. BPN concept is a trivial modification of the mathematical model of neurons.Main feature of BPN is their simplicity, but it has draw back that it converges slowely and lacks of optimality. The first documented description of radial basis functions appeared in the solution of the multivariate interpolation problem. Broomhead and Lowe first exploit the use of RBF in Neural network design in 1988 [32].

In 90's Vapnik and Cortes introduce a new term Support Vector Machine (SVM SVM is a binary learning method that a look for optimal hyperplane as a decision boundary in a high-dimensional space. [1]. SVM Construct a hyperplane in such a way that margin of separation between two classes is maximized [33]. The development of SVM algorithm is the inner dot-product kernel k(x, y) = x, y between a support vector and the input vector.The kernel function produces the dot product in the higher dimensional feature space which reduce execution cost. The improvement in SVM performance depends upon a selection of kernel parameter [34]. Radial basis function is the most popular function among other kernel function [35].

3.2 After 2000's

Mingmin Chi et al. discuss the classification technique for small size training data set. They implement primal SVM for classifying the remote sensing image. Kernel method also plays a central role in designing the primal SVM. The quadratic optimization problem with inequality constraints L_2 norms, SVM uses the Lagrange theory for solving. One of the criticisms of Kernel function is that they cannot produce the unique solution because they are not invertible [36].

Shaohui Tao et al. this paper uses the concept of Least Squares SVM for chemical pattern classification. Tao proposed a fast pruning algorithm for multi output LS-SVM. Principal component analysis or classification correlated component analysis is used in this paper for selecting the support vector data point from the kernel matrix. Then non SV data point transfers the information to the SV data point by transferring method [37].

Tanoori et al. used the concept of active contour method and SVM. This paper authors segment the 3D images of brain for volumetric analysis. First segments brain tissue from intracranial tissues in MR image. Active contour VFC modifies the brain image and extracts the features for classification. SVM classifiers trained for each tissue. The result shows that this algorithm produces a good result as compare to previous approaches [38].

Sugumaran et al. classify the roller bearing faults. They design the automatic diagnostic technique allow unskilled operator to make a decision. For this purpose they extract the feature and select the best feature which is further used as an input to a classification algorithm. They used two methods for classification SVM and PSVM. Observe the result section one can say that PSVM performs steadily better than SVM [39].

Liu et al. proposed a method for segmenting the MRI images. In this paper a feature extraction method is discussed based on a gray symbiotic matrix and gray characteristic values. Then train the SVM for classifying the medical images [40]. YANG et al.classify the apple.it detects the fruit internal quality and then use the classification technique for analyzes information to detect to internal fruit quality. This paper take an X-ray image of fruit and then creating feature vector using SVM and classify the mould core apple and un-mould core apple [41].

Wang et al. proposed algorithm for document classification. It uses the concept of maximum margin projection and least square support vector machine. MMP algorithms used for converting high dimensional data into lower dimensional. LS-SVM is used for classifying the document using sematic rule [42]. Homaeinezhadet.al. proposed ECG heart beat hybrid classification technique. This paper uses the concept of QRS, DWT for feature extraction. Further combined the classification approaches SVM, PNN, MLP for classifying the data set [43].

Lo. et al. classify the breast cancer images using SVM. They detect the breast tissues in MR image and apply SVM for classifying them [44]. JinchangRen Compare two approaches SVM and ANN for detection of breast cancer. SVM perform better than ANN [45].

Yang et al. proposed a method for automatic recognize the poleward moving auroras (PMAs) from all-sky image sequences. Sky image are modeled by HMM and classifying using SVM [46]. Xiang et al. use the color, texture, shape feature used for classification. It proposed multi kernel and multi feature fusion method for image classification. It assigns a weight for each feature. [47].

3.3 Nowadays 2015–17

Nowadays the objective of the research of SVM is focused on their application on practical domain. Image classification is an essential task for image analysis in different areas like environmental studies, military application, hydrological science, agriculture sector, remote sensing etc.

Yuan.et al. proposed a 3D scattering wavelet transform for hyper-spectral image (HIS) classification. It improves the classification performance of classification. The main use of 3-D scattering wavelet transform is that it filter the HIS cube data. Experiment results show that the scattering feature method improves the classification accuracy. This paper has one drawback regarding complexity [48].

Brajesh Kumar et al. introduce an innovative scheme for hyper-spectral image classification. It improves the classification accuracy by using the texture and spectral feature. This method only gives the better accuracy with smaller training set. Moment invariants are used to extract the spatial feature of an image. Texture feature and spatial feature both are used for performing the classification task using SVM [49].

Shruti Goel et al. this paper discusses the new approach for the remote sensing image classification.it uses the artificial intelligent algorithms for improving the classification accuracy. For this meta-heuristic Cuckoo search and Artificial Bee colony algorithm is used. Kappa coefficient value is 0.96 and 0.91. These algorithms use the search space efficiently [50].

Jiangtao Peng et al. use the concept of multi kernel in SVM. It introduces a novel approach region kernel based SVM framework for hyper-spectral image classification. The region to region distance similarity of hyper-spectral image is measured by the Region kernel method. Region kernel based SVM three different types of composite kernels is introduced. The Main focus of this paper is on pixel regions and classifies the local regions [51].

Wen Yang et al. work on semi-supervised feature learning using ensemble projection (EP) and classify the satellite images when very few labeled data is available. First transfer the limited data to an ensemble to WT sets and projecting images onto ensemble WT sets for extracting the preliminary features. For an ensemble of various WT sets, sampling algorithms are used. A new algorithm known as GNA introduces for finding a neighbor in sampling algorithm. Then discriminating functions are successively learned from WT sets [52].

Zebin Wu. et al. used the concept of parallelism for image classification. They parallel implement the composite kernels in SVM for classification task. For designing this framework use the concept of heterogeneous CPU–GPU platform. NVidia's compute unified device architecture (CUDA) programming is used to design a GPU. Composite kernel calculations are done by GPU and rest small data calculation done by CPU. This paper shows that the time complexity of classification is very small and remarkable as compared to conventional computing method. Surely this paper gives the new direction for classification algorithm in terms of computational efficiency and near future this works will explore the other high performance parallel implementations [53].

Yang Song et al. tackle the problem of intra-class variation and inter-class ambiguity in the feature space and resolved by the concept of large margin local estimate classification with sub-category of reference set and apply this concept on medical image classification. Calculate the sub-category of reference set and also generate the local estimates for test image by reference subcategory. Test image classifies by the parameter of similarity with the corresponding class [54].

Yong Luo et al. use the concept of basic mathematics known as matrix completion. Here different features are collected from different views. Multiple features fuses in matrix completion based multi label classification. They combine the weighted MC outputs of different views and developed the multi-view matrix completion framework for classifying the multilevel image [55].

Hongpeng Yin et al. propose a method for scene classification. This paper is use the concept of single layers sparse auto encoder (SAE) for feature learning then applies SVM for classification. Every scene has different complexity from other scene. Extracting a feature from different scene is little difficult. Furthermore, the good classification achieve by this paper with one versus one strategy. We construct S(S-1)/2 SVM classifiers for the data set which contain S scene category. This paper presents a good result only if S scene is a small number of as compare large [56].

Katta Sugamya et al. proposed a work for feature extraction at a lower level and classify with SVM. It uses wavelet transformation for shape feature, color correlogram for color feature and Gabor wavelet for texture feature. Farther combines these features and find similarity using the SVM classifier. If we want to increase the robustness of this method then investigate the most effective method for extracting the feature [57].

Shuming Liu et al. SVM use for classifying the smiley face and use Candide-3 model for selecting the characteristic parameter of the face. This paper work on 3D facial image and extract facial feature eyes, eyebrows, lips, mouth and other low level local feature. Apply SVM for classifying and recognize the facial expression [58].

Oliva et al. This paper is colonoscopy image for identifying the diseases and predict as well classify the diseases. They evaluate colon tissue images, extract the features, then classify the image normal and abnormal. They use the MIAS open source technology for improving their flexibility and maintainability. The main approach for this is a first load image and selects the fragment of the image, then extracts the feature and applies this feature to WEKA. Load the classification model built with WEKA. The Result shows that MIAS is able to predict and classify the image more accurate than other method [59].

Zijun Wei et al. for image classification, author uses the concept of pooling local information. Pooling method has two common approaches known as average pooling and max pooling. They propose the Region Ranking SVM. They give the rank to each region by subsequence score distribution. It learns a region evolution function and combining region [60].

Jie Ding et al. classify the tongue images using SVM. The main problem for a tongue image data set is an extraction of appropriate feature and robust approach for classification. First extract the local feature of tongue image. These features selected by HOG concept and apply this feature into classifiers for prediction. SVM and doublet concept have good classification accuracy. This work can be extended for more data set. This method includes the following steps. First extract the feature using HOG method. Secondly, build a sample for Doublet. Then calculate the distance metrics by the Doublet SVM classifier and predict the disease. Experiment result shows good classification and prediction accuracy [61].

Sachdeva et al. classify and analysis brain tumors. Genetic approach is applied for selecting the optimal feature. Content based active region method is marked the tumor regions. Intensity and texture feature are selected from the input image data set. Genetic algorithm uses the probabilistic approach for identifying the tumor class. The result shows that SVM increase the classification accuracy 80.8–89% [62].

Imène Garali et al. propose the new diagnostic approach for classification of medical image. They use the Positron Emission Tomography Brain Images for prediction of neurodegenerative diseases. First segment the image for selecting volumes of interested (VOI). Calculate the first and second derivative VOI. Applying these values as input to the SVM classifiers, yielding the better classification accuracy applying the intensity mean value as a feature [63].

S. Park et al. propose the scheme for a seed growing framework to improve interactive segmentation. The geodesic distance feature is used input to SVM classifier. An SVM classifier trained with seed super pixel of an input image. Experiment result shows that proposed method are improving the segmentation accuracy. Future scopes of this frame work will extended on large database and resolve the seed bias problem [64].

Wang et al. explore the use of one class support vector machine for selection hyper-parameter using self-adaptive data shifting methodology. This scheme can efficiently create high-quality pseudo outlier and target data to assess the error on outlier class and target class respectively, without acquainting any new hyper-parameters with be tuned by users [65]. Mahesh et al. classify the gender on the basis of facial images. Zernike Moments based shape descriptors are used for classifying the images and different classifiers are used to test the accuracy of the results [66].

Priya et al. proposed a method for plant classification on the basis of leaf Recognition. In this paper different features are extracted from leaf images then apply PCA for minimize the dimension of values. These feature values act as input to SVM for classification [67].Es-saady et al. use the concept of SVM for identifying plant leaves diseases. Two serial SVM classifiers are used for diseases recognition [68]. Tomar et al. work on directed acyclic graph based multi-class least square twin support vector machine for classifying plants by leaf recognition. Hybrid feature selection (HFS) approach is utilized to get the best discriminant features for the recognition of individual plant species [69]. Lukic et al. also work on leaf recognition for plant classification. SVM are used as classifier and Hu moments uniform local binary pattern histogram parameters as features [70].

4 Conclusion

Image classification has made countless growth over past decade. This paper summarizes the comprehensive overview with generalized the typical functionality and highlighting their application in remote sensing and medical field. The main contribution of this paper is briefly reviewing the concept of SVM and SVM classification accuracy in the different data set. It is well known that classification methodologies may vary with different type image data. This paper mainly focuses on SVM method during 2015–2017.

This study shows that there are many ways of modifying the method to get better classification accuracy. Accuracy assessment is an integral part in image classification technique. Features are contained the informative data for image classification. Extraction and Selection of informative, feature play an important role in achieving the accuracy. So this paper also discusses the various feature selection method and kernel parameter selection strategies.

The success of an image classification depends on many factors like high quality image data set, ancillary data, extracted feature, feature selection and designing an appropriate classification method. A relative study of different classification method is helpful for enhancement of classification accuracy. It is essential for the forthcoming investigation to develop guidelines on the applicability and capability of major classification methodology. This survey of different technique provides a platform for the designing a new novel scheme in this area as a future work.

References

1. Cortes C, and Vapnik V (1995) Support-vector network. Mach Learn 20(3):273–297
2. Tomar D, Agarwal S (2015) A comparison on multi-class classification methods based on least squares twin support vector machine. Knowl-Based Syst 81:131–147
3. Mangasarian OL, Musicant DR (2001) Lagrangian support vector machines. J Mach Learn Res 1:161–177
4. Xiang Z, XueqiangLv, Zhang K (2014) An Image Classification Method Based On Multi-feature Fusion and Multi-kernel SVM. In: Seventh International Symposium on Computational Intelligence and Design, Hangzhou, p 49–52
5. Hsu C-W, Lin C-J (2012) A comparison of methods for multi-class support vector machines. IEEE Trans Neural Networks 13(2):415–425
6. Milgram J, Cheriet M, Sabourin R (2006) One Against One or One Against All: Which One is Better for Handwriting Recognition with SVMs. In: Tenth International Workshop on Frontiers

in Handwriting Recognition, Oct 2006, La Baule (France), Suvisoft,
7. Jayadeva R, Khemchandani R, Chandra S (2007) Twin support vectormachine for pattern classification. IEEE Trans Pattern Anal Mach Intell 29(5):905–910
8. Han J, Kamber M (2001) Data mining: concepts and techniques. Morgan Kaufmann Publishers, USA
9. Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46:389–422
10. Prajapati GL,and Patle A(2010) On Performing Classification Using SVM with Radial Basis and Polynomial Kernel Functions. In: Third International Conference on Emerging Trends in Engineering and Technology, p 512–515
11. Kuo B-C, Ho H-H, Li C-H, Hung C-C, Taur J-S (2014) A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. IEEE J sel Topic Appl Earth Observ Remote Sens 7(1):317–326
12. Morariu D, Vintan LN, Volker Tresp V (2006) Feature selection methods for an improved SVM classifier. Trans Eng 14:83–89
13. Huang C, Chen M, Wang C (2007) Credit scoring with a data mining approach based on SVMs. National Kaohsiung First university of Science and Technology, Department of Information Management, Nantz District, pp 847–856
14. Lin W, Lee Z, Lee S (2008) Parameter determination of support vector machine and feature selection using simulated annealing approach. Applied Soft Comput 8:1505–1512
15. Wang Z, Zhu C, Niu Z, Gao D, Feng X (2014) Multi-kernel classification machine with reduced complexity. Knowl-Based Syst 65:83–95
16. Scholkopf B, Burges CJC, Smola AJ (1999) Advances in kernel methods: support vector learning. MIT Press, Cambridge
17. Du P, Tan K, Xing X (2010) Wavelet SVM in reproducing kernel hilbert space for hyper-spectral remote sensing image classification. Opt Commun 283(24):4978–4984
18. Haar A (1910) Zurtheorie der orthogonalen funktionen systeme. Math Ann 69:331–371
19. Graps A (1995) An Introduction to Wavelets. IEEE Comput Sci Eng 2(2):50–61
20. Agarwal E, Gupta S, Chandra MA (2014) Data hiding using lazy wavelet transform strategy. IJCA Proceedings on International Conference on Advances in Computer Engineering and Applications. ICACEA 5:5–8
21. Wickerhauser MV (1992) Acoustic signal compression with wave packets: wavelets: a tutorial in theory and applications. Academic Press Professional, San Diego, pp 679–700
22. Myint SW, Zhu T, Zheng B (2015) A novel image classification algorithm using over complete wavelet transforms. IEEE Geosci Remote Sens Lett 12(6):1232–1236
23. Nguyen T, Khosravi A, Creighton D, Nahavandi S (2015) Classification of healthcare data using genetic fuzzy logic system and wavelets. Expert Syst Appl 42:2184–2197
24. Nguyen T, Khosravi A, Creighton D, Nahavandi S (2015) Medical data classification using interval type-2 fuzzy logic system and wavelets. Applied Soft Comput 30:812–822
25. Tan Y, Li G, Duan H, Li C (2014) Enhancement of medical image details via wavelet homomorphic filtering transform. J Intell Syst 23(1):83–94
26. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133
27. Rosenblatt F (1957) The perceptron—a perceiving and recognizing automaton. Cornell aeronautical laboratory, report number: 85-460-1
28. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69
29. Broomhead D S, lowe D (1988) Radial basis functions, multi-variable functional interpolation and adaptive networks, royal signals and radar establishment memorandum No. 4148, 38
30. Widrow B, Hoff ME (1960) Adaptive switching circuits. IRE WESCON Conven Rec 4:96–104
31. Rumelhart DE, McClelland JL (eds) (1986) Parallel distributed processing: explorations in the microstructure of cognition, vol 1. MIT Press, Cambridge
32. Broomhead DS, Lowe D (1988) Multivariable function interpolation and adaptive networks. Complex Syst 2:321–355
33. Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V (2000) Feature Selection for SVMs. In: Proceeding NIPS'00 Proceedings of the 13th International Conference on Neural Information Processing Systems. p 647–653
34. Chapelle O, Vapnik V, Bousquet O, Mukhetjee S (2000) Choosing kernel parameters for support vector machines. Mach Learn 46(1–3):131–159
35. Burbidge R, Buxton B (2001) An introduction to support vector machines for data mining. Computer Science Dept.,UCL, UK
36. Chi M, Feng R, Bruzzone L (2008) Classification of hyper-spectral remote-sensing data with primal SVM for small-sized training dataset problem. Adv Space Res 41:1793–1799
37. Tao S, Chen D, Zhao W (2009) pruning algorithm for multi-output LS-SVM and its application in chemical pattern classification. Chemometr Intell Lab Syst 96:63–69
38. Tanoori B, Azimifar Z, Shakibafar A, Katebi S (2011) Brain volumetry: an active contour model-based segmentation followed by SVM-based classification. Comput Biol Med 41:619–632
39. Sugumaran V, Ramchandran KI (2011) Effect of number of features on classification of roller bearing faults using SVM and PSVM. Expert Syst Appl 38:4088–4096
40. Liu Y-T, Zhang H-X, Li P-H (2011) Research on SVM-based MRI image segmentation. J China Univ Posts Telecommun 18(2):129–132
41. YANG L, YANG F, NOGUCHI N (2011) Apple Internal Quality Classification Using X-ray and SVM. In: Proceedings of the 18th World Congress The International Federation of Automatic Control IFAC Proceedings, Volume 44, Issue 1, January 2011, p 14145–14150 Milano (Italy)
42. Wang Z, Sun X (2011) Document classification algorithm based on MMP and LS-SVM. Advanced in Control Engineering and Information Science. Procedia Eng 15:1565–1569
43. Homaeinezhad MR, Tavakkoli E, Atyabi SA, Ghaffari A, Ebrahimpour R (2011) Synthesis of multiple-type classification algorithms for robust heart rhythm type recognition: neuro-svm-pnn learning machine with virtual QRS image-based geometrical features. Scientia Iranica B 18(3):423–431
44. Lo C-S, Wang C-M (2012) Support vector machine for breast MR image classification. Comput Math Appl 64:1153–1162
45. Ren J (2012) ANN vs. SVM: which one performs better in classification of MCCs in mammogram imaging. Knowl-Based Syst 26:144–153
46. Yang Q, Liang J, Hub Z, Xing Z, Zhao H (2012) Automatic recognition of poleward moving auroras from all-sky image sequences based on HMM and SVM. Planet Space Sci 69(2012):40–48
47. Xiang Z, Lv X, Zhang K (2014) An Image Classification method based on multi-feature fusion and multi-kernel SVM. In: 2014 seventh international symposium on computational intelligence and design, Hangzhou, p 49–52
48. Tang YY, Lu Y, Yuan H (2015) Hyper-spectral image classification based on three-dimensional scattering wavelet transform. IEEE Trans Geosci Remote Sens 53(5):2467–2480
49. Kumar B, Dikshit O (2015) Spectral-Spatial Classification of Hyper-spectral Imagery Based on Moment Invariants. IEEE

Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(6):2457–2463

50. Goela S, Gaur M, Jain E (2015) Nature inspired algorithms in remote sensing image classification. Procedia Comput Sci 57:377–384
51. Peng J, Zhou Y, Chen CLP (2015) Region-kernel-based support vector machines for hyper-spectral image classification. IEEE Trans Geosci Remote Sens 53(9):4810–4824
52. Yang W, Yin X, Xia G-S (2015) Learning high-level features for satellite image classification with limited labeled samples. IEEE Trans Geosci Remote Sens 53(8):4472–4482
53. Wu Z, Liu J, Plaza A, Li J, Wei Z (2015) GPU implementation of composite kernels for hyper-spectral image classification. IEEE Geosci Remote Sens Lett 12(9):1973–1977
54. Song Y, Cai W, Huang H, Zhou Y, Feng DD, Wang Y, Fulham MJ, Chen M (2015) large margin local estimate with applications to medical image classification. IEEE Trans Med Imaging 34(6):1362–1377
55. Luo Y, Liu T, Tao D, Xu C (2015) Multiview matrix completion for multilabelimage classification. IEEE Trans Image Process 24(8):2355–2368
56. Yin H, Jiao X, Chai Y, Fang B (2015) Scene classification based on single-layer SAE and SVM. Expert Syst Appl 42:3368–3380
57. Sugamya K, Pabboju S, Babu S V (2016) A CBIR Classification Using Support Vector Machines. In: International Conference on Advances in Human Machine Interaction (HMI—2016)
58. Liu S, Chen X, Fan D, Chen X, Meng F, Huang Q (2016) 3D Smiling Facial Expression Recognition Based on SVM. In: 2016 IEEE International Conference on Mechatronics and Automation, Harbin, p 1661–1666
59. Oliva J T, Lee HD, Spolaôr N, Coy CSR, Wu FC (2016) Prototype system for feature extraction, classification and study of medical images. Expert Syst Appl 63 No. C, 267–283
60. Wei Z, Hoai M (2016) Region Ranking SVM for Image Classification. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, p 2987–2996
61. Ding J, Cao G, Meng D (2016) Classification of Tongue Images Based on Doublet SVM. In: International Symposium on System and Software Reliability (ISSSR), Shanghai, p 77–81
62. Sachdeva J, Kumar V, Gupta I, Khandelwal N, Ahuja CK (2016) A package-SFERCB-Segmentation, feature extraction, reduction and classification analysis by both SVM and ANN for brain tumors. Appl Soft Comput 47:151–167
63. Garali I, Adel M, Bourennane S, Guedj E (2016) Classification of positron emission tomography brain images using first and second derivative features. In: 2016 6th European Workshop on Visual Information Processing (EUVIP), Marseille, pp. 1-5.EUVIP 201, p 1–5
64. Park S, Lee HS, Kim J (2017) Seed growing for interactive image segmentation using SVM classification with geodesic distance. Electron Lett 53(1):22–24
65. Wang S, Liu Q, Zhu E, Porikli F, Yin J (2018) Hyperparameter selection of one-class support vector machine by self-adaptive data shifting. Pattern Recogn 74:198–211
66. Mahesh VGV, Raj ANJ (2018) Zernike moments and machine learning based gender classification using facial images. In: Abraham A, Cherukuri A, Madureira A, Muda A (eds) Proceedings of the Eighth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2016). SoCPaR 2016. Advances in Intelligent Systems and Computing, vol 614. Springer, Cham
67. Priya C A, Balasaravanan T, Thanamani A S (2012) An efficient leaf recognition algorithm for plant classification using support vector machine. In: International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012), Salem, Tamilnadu, p 428–432
68. Es-saady Y, Massi I El, Yassa I El, Mammass, D, A. Benazoun A (2016) Automatic recognition of plant leaves diseases based on serial combination of two SVM classifiers. In: 2016 International Conference on Electrical and Information Technologies (ICEIT), Tangiers, p 561–566
69. Tomar D, Agarwal S (2016) Leaf recognition for plant classification using direct acyclic graph based multi-class least squares twin support vector machine. Int J Image Graph 16(3):1650012
70. Lukic M, Tuba E, Tuba M (2017) Leaf recognition algorithm using support vector machine with Hu moments and local binary patterns. In: IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any p 000485–000490