Computer Science and Information Technologies
Vol. 4, No. 3, November 2023, pp. 226~239
ISSN: 2722-3221, DOI: 10.11591/csit.v4i3.pp226-239

Generalization of linear and non-linear support vector machine in multiple fields: a review

Sundas Naqeeb Khan[1], Samra Urooj Khan[2], Hanane Aznaoui[3], Canan Batur Şahin[4], Özlem Batur Dinler[5]
[1]Department of Graphics, Computer Vision, and Digital Systems, Silesian University of Technology, Gliwice, Poland
[2]Department of Electrical Engineering Technology, Punjab University of Technology, Rasul, Mandi Bahauddin, Punjab, Pakistan
[3]Faculty of Computer Science, Cadi Ayyad University, Marrakesh, Morocco
[4]Faculty of Computer Science, Turgut Özal University, Ankara, Turkey
[5]Faculty of Computer Science, Siirt Üniversitesi, Siirt, Turkey

Article Info

Article history:

Received Mar 22, 2023
Revised May 28, 2023
Accepted Jun 29, 2023

Keywords:

Sentiment analysis
Generalization
Oversimplification
Pre-processing
Support vector machines

ABSTRACT

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. In other terms, SVM is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy. In this article, the discussion about linear and non-linear SVM classifiers with their functions and parameters is investigated. Due to the equality type of constraints in the formulation, the solution follows from solving a set of linear equations. Besides this, if the under-consideration problem is in the form of a non-linear case, then the problem must convert into linear separable form with the help of kernel trick and solve it according to the methods. Some important algorithms related to sentimental work are also presented in this paper. Generalization of the formulation of linear and non-linear SVMs is also open in this article. In the final section of this paper, the different modified sections of SVM are discussed which are modified by different research for different purposes.

Corresponding Author:

Sundas Naqeeb Khan
Department of Graphics, Computer vision and digital systems, Silesian University of Technology
Akademicka Street 2A, 44-100 Gliwice, Poland
Email: sndskhan87@gmail.com

1. INTRODUCTION

Oversimplification performance for pattern recognition with any classifier can be attained while the classification capability is achieved along with a better formulation function. This formulation function must apply to the training set for the matching process in which the size is the most favorable factor. Instead of the functions, classifiers keep several parameters for doing work position but these parameters change according to the situation, so these parameters have moldability, due to this, known as amendable parameters. Based on classification capability, these parameters be trained quickly about the training set devoid of any error but display a slightly low simplification method [1]. On the contrary, a classifier with non-amendable parameters or functions can't maintain its capability for competent learning at all. Between them, the classifier's finest capability for competence adjustment minimizes the estimated simplification error for the training set which is provided as an original form. According to the empirical proof with theoretical investigation associated with the simplification error for the training set with the help of a classifier, the classifier also resolves the complexity issue in learning [2], [3]. Risk minimization error becomes the foundation of the unique classifier that sustains the ability to solve this issue with very high performance for pattern recognition named support vector machines (SVMs).

Nowadays this classifier solves a large number of problems in regression and classification sections with high accuracy. The concerning point of SVMs is to explore a hyper-plane within n-dimensional space which has particularly categorized the data according to its own selected data points called support vectors [4]. SVM belongs to the category of supervised learning that is linked with learning algorithms in which analysis of data provides accuracy for classification and regression problems. SVM is a representative of training sets (data points) in a bounded region, separated into two classes of data with a reasonable distance with the help of support vectors [5]. Thus, with the separation process of SVM, the bounded region is divided into two marginal categories with suitable distance while the gap between two separable classes, and support vectors should be as wide as possible. The main reason behind the wide range of distance between selective support vectors is new prediction [6].

SVM performs two types of classification such as linear and non-linear. Linear classification is active when data of classes are in labeled form. On the other hand, supervised learning cannot handle the non-linear problem of classification. On behalf of this, researchers used a trick which is known as a kernel in which the mapping process shifts their inputs into high dimensional feature space, to solve these types of problems [7]. On the record, an SVM constructs a hyper-plane into high dimensional space in which separated regions with decided boundaries become the reason for the largest distance between selected support vectors. The gap between the selected support vectors of two classes is called functional margin. In general, when the margin is low then the simplification error of the classifier is also low vice versa [8]. The primary issue can be in a predetermined dimensional space in which the data sets are not present in the form of a linear set, due to this reason, this space turns into high dimensional space. Reasonable, the mapping process of SVM operates a dot product function in which the input data appears in pairs. As a subject, these pair vectors can be calculated in terms of variables with their kernel function $\mathrm{k}(x, y)$. With the help of a conversion scheme, high dimensional space consists of data sets and the points of including these data sets perform dot product along a vector such as a set of vectors is minimal. The vectors satisfy the pre-defined hyper-planes with their linear parameters $\propto_i$ of classes of feature vector x_i database. A point x in the feature space which is going to be mapped on hyper-plane, describes the relationship of $\sum_i^n \propto_i k(x_i, x) = constant$,where $k(x, y)$ become small y extended as compared to x, every measurement of summation is grown to be the degree of proximity of test point x to the equivalent database x_i. Besides this, kernel summation is used for the measurement of the comparative proximity of every test point while this function construct or designs the optimal solution for the relative problem with some modifications [9]. The usage of SVM can vary according to the domain of the problem and solve a variety of real-world problems such as text categorization, image classification, recognition of handwritten characters, and bioinformatics [10].

2. RECENT LITERATURE

This section emphasizes the most related work for other researchers who used SVM for the generalization of their research problems. Here we describe some significant research portions that are based on the algorithms. The SVM algorithm is used for classifying the data according to the steps in Figure 1.

Input: I: Input data
Output: V: Set of support vectors
Begin
Step 1: Divide the given dataset into two set of data items having different class labels assigned to them
Step 2: Add them to support vector set V
Step 3: Loop the divided *n* data items
Step 4: If a data item is not assigned any of the class labels, then add it to set V
Step 5: Break if insufficient data items are found
Step 6: End loop
Step 7: Train using the derived SVM classifier model and test so as to validate over the unlabeled data items
End

Figure 1. Steps of SVM for sentiment analysis

According to Pavlidis *et al.* [11] research is based on the performance improvement in sentiment analysis regarding classification. The proposed methodology presents the combination of SVM and Naïve Bayes (NB) along with excellent results concerning accuracy metrics. They give two new algorithms for sentiment analysis according to the word and sentence level and these algorithms [12] perform preprocessing steps on selected datasets and translate unstructured forms of reviews into structured forms furthermore structured forms translate into numerical values with the help of the lexicon scoring method. Therefore, the focused area of research in this research is feature selection and semantic analysis. For classification purposes, SVM is used with its radial basis kernel function. Whereas, a dataset $D = \{X_i, y_i\}$ as X_i present the set of records with class labels y_i. A separating hyper-plane is used $W^{*}x + b = 0$, where, $W = \{w_1, w_2, \ldots, w_n\}$ as w_n is a weight vector with n features while b shows biases [13]. To achieve the goal, the maximum margin hyper-plane with the help of lagrangian calculation in terms of formula for record testing as: in SVM classifier perform operations for accuracy according to the $f(x) = \sum_{i=1}^{n} \propto_i k(X, X_i) + b$ given formula, where $k(X, X_i)$ is the radial basis kernel function along $exp\left(-\frac{\|X-\acute{X}\|^2}{2\sigma}\right)$ formulation. Another research article presented text classification through cosine similarity with latent semantic indexing for the Arabic language. In the reference article [14], the proposed algorithm is based on some useful steps. Table 1 illustrate the related literature about the usage of SVM with its properties, functions, datasets, algorithm modifications, rules, and hybridization to solve different types of problem in the real world [15].

Table 1. Different research fields where SVM used for excellent results

Research areas	Count frequency
Sentiment analysis	[16]–[20]
Text categorization	[21]–[23]
Image classification	[24]–[26]
Bio-informatics	[27]–[30]
Other fields	[31]–[33]

3. BACKGROUND INFORMATION ABOUT SUPPORT VECTOR MACHINES

In the past two decades, machine learning progressed in the knowledge domain and performs actions as a backbone. It evolves as a pivot point in our daily life. In the word world, data is growing day by day then control of the access process to the data availability turns into a problem. The logical motivation to consider is that elegant data analysis should be inescapable from the technical development. Machine learning divided the outer factors into two major fields such as art and science. In the art field, it tries to decrease some contrasting category issues with the help of quite contracted models while in the science field, providing solutions to the issues with generalization performance [34], [35]. Intellectual learning is a field of processes that describe complicated tasks specifically concerning understandable methods. According to the standard definition of learning is “to increase knowledge with high indulgent or expertise through, learning, training, and practice”. Therefore, learning about machine systems, mechanisms, infrastructure, approaches, and working styles is collectively known as machine learning. Normally, the modifications of a system are submitted to the work positions to operate several activities directly connected to artificial intelligence. Several categories of activities engross in detection, analysis, arrangement, robot control systems, and calculation. The modifications can be relevant to the enhancements or can be new [36].

The main conceptual idea about machine learning work can be comprehended in the light of four broader perspectives. Firstly, its focal point is prediction while the prediction keeps some background knowledge about things. Therefore, that background knowledge is useful in prediction such as based on the past, and what they did. Secondly, it searches out the relationship between original data and prediction practically. Thirdly, it can exploit a set of stimuli/inputs for prediction with statistical approaches in technical form. Lastly, it tries to predict the value of a variable Y that is given an input of feature set X [37]. Machine learning is divided into three main categories such as supervised learning, semi-supervised learning, and unsupervised learning furthermore, these categories sub-divisions are shown in Figure 2. The study of the support vector knowledge is beneficial from two major perspectives. Firstly, the theoretical part is strong enough for satisfactory analysis in which the work foundation is pleasant with straightforward initiatives and designs with a strong structure. Secondly, the guidance about any experimental application with high accomplishment achieved. The intersection between the theoretical portion with experimental support vectors becomes an intelligence of work for practical life [38]. Several categories of algorithms along arithmetical tests might be classifying knowledge in a particular field and environmental factors support its structure. However, the real world needs to use and learn the most multifaceted designs and algorithms just like neural networks which are more complex than support vectors in the form of theoretical analysis [39].

The SVM predicts the supporting vectors for the separation of two classes with high margins. Support vectors perform mathematical operations easily due to their special correspondence regarding a linear scheme in a high-dimensional feature space. This feature space is non-linearly associated directly with input space. Furthermore, a linear structure in a high dimensional space performs accurate operations in experimental applications but the formulation of any mathematical equations cannot be computed in that structural space. Therefore, kernels are a type of trick for solving this issue of computation [40]. Kernels achieve any type of formulation with evidence straight in that input structural space. A very nice solution provided by a curl of support vectors is kernels.

Nowadays, researchers focus on non-linear trends for detection, deterioration problems, or mining of significant features of useful applications. Support vectors become the root of some branches such as the theoretical part of any knowledge, to get the most favorable results for hyper-plane algorithms, the computation of kernel operations, and its functional evaluation [41]. Hyper-planes: according to the mathematical term, a hyper-plane is divided into sub-spaces where the measurement belongs to its circulating spaces. There are some specific learning rules for describing the relationship between the hyper-plane and its spaces: if a space is represented in 3D it means its relation with the hyper-plane is in 2D planes while if the space performing its operation within 2D then its associated hyper-plane is in 1D line. These rules are implemented normally in generic spaces which belong to the idea about the structure of the sub-spaces. Hyper-planes help to support vectors for their activities regarding pattern recognition or natural language. Figure 3 illustrates many unique categories of spaces used in different types of hyper-planes. Every space has some sub-category functions e.g., branch space has norm and completeness as sub-functions [42]. Therefore, the presented result of this hyper-plane is going to be a solution of a single linear mathematical form such as $a_1x_1 + a_2x_2 + \ldots + a_nx_n + b = 0$. Projective hyper-plane extended the concept of the plane. Sub-division of the spaces is not possible in this form but it considers the two hyper-planes for division of the points and then space. While it keeps a set of points along its property. For any two points from the set construct a learning rule for examination of the rest of all points in the set e.g., a lone hyper-plane in which all the sides are connecte [43], [44]. For the structure of the algorithm, a combination of functions is required for a class in which capacity may be involved as a primary objective. Support vectors become the foundation of hyper-planes equivalent towards the conclusion purpose $(\mathrm{w.x}) + \mathrm{b} = 0$ where $\mathrm{w} \in \mathrm{R}^{\mathrm{N}}$, $\mathrm{b} \in \mathrm{R}$. This is a simple equation for getting optimal hyper-plane within space [40]. Generalization of SVM: machine learning is a sub-field of artificial intelligence in which the main research is about the expansion phase of the approaches with their suitable schemes whereas the computer learns and is used as a machine. Moreover, the expansion phase included algorithms in which the machine learning process is considered its major task with their sub-activities. In so many traditions, machine learning has common characteristics among statistical progress. With time, various useful approaches with their structural design proposed for machine learning activities [45].

The history of the SVM algorithm originated in 1963 by Vapnik and in 1992, the first time SVM was introduced based on a set of correlated supervised learning methods, and these methods were used for classification and regression problems as non-linear classifiers. According to the non-linear classifier, Vapnik used kernel tricks for getting maximum margin hyper-planes [46], [47]. SVM is a specific type of classification while the working layout depends on the conversion process where the training dataset is transformed into a higher dimensional space in which the main purpose is to investigate how to divide decision boundaries among classes. According to these boundaries, hyper-planes become useful boundaries in which support vectors can be classified rather than other data points with their margin. These margins are present in the form of parallel lines which describe through the shortest distance between a hyper-plane and its support vectors. Therefore, SVM is capable of classification of both types of datasets such as linear and non-linear [48], [49]. SVM roughly sketched for some significant issues [50] that is: class division: researchers tried to obtain the optimum extrication solution for hyper-plane among the selected classes through maximizing the distance. The distance between decision boundaries and support vectors is called margin as shown in Figure 3 and in between these margins, researchers get their optimal hyper-plane. Internal representation of classes: selected support vectors reside in the opposite direction of the prominent distance effect on the margin and reduce its effect in terms of weight. Non-linear property of classes: there is a big issue regarding if the linear separation cannot be found then selected support vectors can expected towards higher dimensional space whereas these selected data points used kernel trick while efficiently participating in linearly separable operation [51]. Weights of classes: this depends on the particular vector which is known as the weight component such as A and B are two different classes with asymmetric class sizes with weight vectors. Cross-validation for classification: for training data evaluation, k-fold cross-validation is executed through different possible combinations of the parameters in which most of the time default values are set for good accuracy.

Why does the SVM margin is $\frac{\mathbf{2}}{\|\mathbf{w}\|}$: by using geometry, the margin is probably related to the maximum distance between two parallel hyper-planes that can separate the two-point sets; let x_0 be a point in the hyper-plane of $\mathrm{wx} - \mathrm{b} = -1$, so this hyper-plane equation becomes $\mathrm{wx_0} - \mathrm{b} = -1$ [52], [53]. To measure the distance between hyper-planes $\mathrm{wx} - \mathrm{b} = -1$, and $\mathrm{wx} - \mathrm{b} = +1$, there is only a need to compute the perpendicular distance from x_0 to plane $\mathrm{wx} - \mathrm{b} = 1$, denoted as r. Therefore, $\frac{\mathrm{w}}{\|\mathrm{w}\|}$ is a unit normal vector of the hyper-plane $\mathrm{wx} - \mathrm{b} = 1 \Rightarrow \mathrm{w}\left(\mathrm{x_0} + \mathrm{r}\frac{\mathrm{w}}{\|\mathrm{w}\|}\right) - \mathrm{b} = 1$, where $\left(\mathrm{x_0} + \mathrm{r}\frac{\mathrm{w}}{\|\mathrm{w}\|}\right)$ should be a point in hyper-plane, according to the definition of r, equation of hyper-plane becomes as [54]. General example: an easy example selected from our daily life. Let's suppose here a scenario-1 is given for the testing approach. Scenario-1: "I have a business and I receive a lot of emails from customers every day. Some of these emails are complaints and should be answered very quickly. I would like a way to identify them quickly so that I answer these emails in priority". According to this given scenario [55], we try to provide the best optimum solution with the help of a supervised machine learning algorithm with some significant steps. Now a supervised learning algorithm e.g., SVM is used for training sessions along the labeled dataset as a linear model. So in simple words, the linear model is based on a line that separates the data hyper-plane. Some points are described in Figure 4 and these play an important role in the process of separating the line in hyper-plane [56], [57].

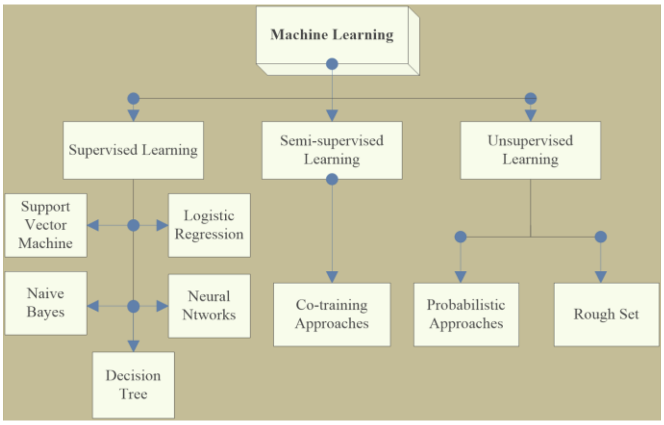

Figure 2. Sub-division of machine learning approaches

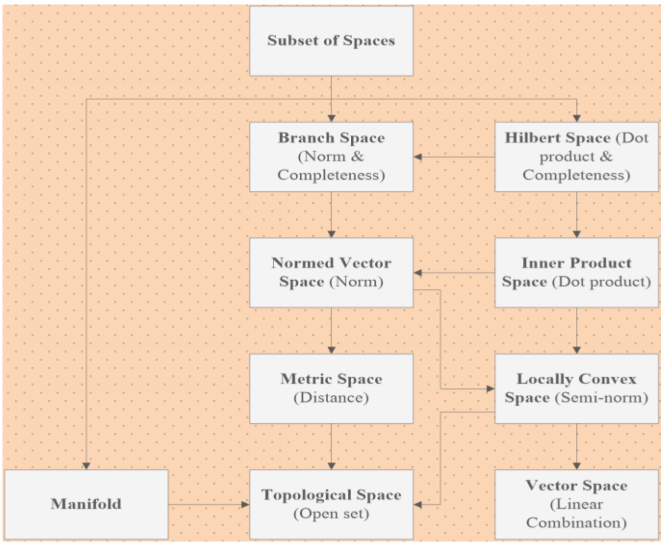

Figure 3. Types of abstract spaces of hyper-planes

Step-1: Basic need for a lot of emails, the more the better.

Step-2: Ready the title of each email and classify it by saying "it is a complaint'' or "it is not a complaint''. It put a label on each email.

Step-3: Train a model on this dataset.

Step-4: Assess the quality of the prediction (using cross validation).

Step-5: This mode is used to predict if an email is complaint of not.

Figure 4. Learning properties for a line with the separation process

There are four major types of SVM used for classification tasks in any field of research. In terms of scenario 1, SVM is divided into four categories of classification: i) the original one: the maximum margin classifier; ii) the kernelized version using the kernel trick; iii) the soft margin version; and iv) the soft margin kernelized version in which i), ii), and iii) are combined.

These four types merged into two types of SVM to linear and non-linear SVM. Figure 4 illustrates the linear SVM in the best way. Formulation of SVM problem: In this type of SVM, the classifier basic element is a separating hyper-plane with the line equation along [+1 -1] interval. The significant points that belong to the training sets are considered support vectors which these points have the properties of the separating process of hyper-plane. Mathematically, a linear SVM-focused equation is the separation of the hyper-plane with its optimum line as (1) to (4) [58].

$$W.X^{+} + b = +1 \tag{1}$$

$$W.X^{-} + b = -1 \tag{2}$$

$$W.(X^{+} - X^{-}) = 2 \tag{3}$$

$$M = \frac{(x^{+} - x^{-}).w}{|w|} = \frac{2}{|w|} \tag{4}$$

Where +1 belongs to the positive predicted class, -1 belongs to the negative predicted class, *w* is the weight vector and M is the margin width.

Consequently, the main achievement through linear SVM can be obtained by the formulation according to the formulas along their properties as (5) [59].

$$\left(\begin{cases} wx_i + b \geq 1 \; if \; y_i = +1 \\ wx_i + b \leq 1 \; if \; y_i = -1 \\ y_i(wx_i + b) \geq 1 \; if \;\; \forall_i \end{cases}\right) \tag{5}$$

To maximize the margin $\frac{2}{|w|}$, and minimize $\frac{1}{2}w^T w$, researchers formulate a quadratic optimization problem through the given equation with its subject constraint as (6).

$$\{minimize \; \emptyset(w) - 1/2 \; w\hat{}t \; w \; subject \; to \; y\_i \; (wx\_i + b) \geq 1 \; \forall\_i\} \tag{6}$$

For the proper solution, to solve the optimization problem, to find this statement: $\left\{(w \; and \; b) | \min \emptyset(w) = \frac{1}{2} W^T W \; and \; \forall \; \{(X_i, Y_i)\}: y_i(W^T X_i + b) \geq 1\right\}$ [60]. Therefore, for the optimization of a quadratic function, linear constraints become a powerful subject. In mathematical programming problems, quadratic optimization problems are a very famous category for optimal solutions. These solutions include the structural design which is known as a dual problem in which a Lagrange multiplier named $\propto_i$ is linked along each and every constraint within the primary problem [61]. For solving the optimization problem, some properties combined in a formulation as (7):

$$Find \; (\propto_1, \ldots, \propto_N) | \max \boldsymbol{Q}(\propto) = \sum \propto_i - \frac{1}{2} \sum \sum \propto_i \propto_j \; y_i y_j X_i^T X_j \; subject \; to \; \sum \propto_i \; y_i = 0, \propto_i \geq 0 \; \forall \propto_i \tag{7}$$

the formation of the solution depends on *w* and *b* as (8):

$$\{(w = \sum \propto_i \; y_i X_i) \; and \; (b = y_k - W^T X_k) \; for \; any \; X_k | \propto_k \neq 0\} \tag{8}$$

whereas each $\propto_i$ shows that it corresponding X_i. This will be a support vector. Based on the category of problem, the classifying function will have the form (9) [62].

$$f(X) = \sum \propto_i \; y_i X_i^T + b \tag{9}$$

This function depends on, an inner product between the testing point *x* and its corresponding sup isport vector x_i. Calculation of the inner products $X_i^T X_j$ between all pairs of points of the training set involved in the solution of the optimization problem. As per instructions, if noise detects in the training set then slack variables ξ_i can be added in the function and allow misclassification of difficult data. After adding the slack variable in the quadratic optimization problem, the (10) will become next, and the representation is shown in Figure 5.

$$\left\{\min \emptyset(x) = \frac{1}{2} W^T . W + C \sum_{k=1}^{N} \xi_k \ \forall\{(X_i, y_i)\} | y_i(W^T X_i + b) \geq 1 - \xi_i \ and \ \xi_i \geq 0 \ \forall_i\right\} \quad (10)$$

Where C indicates the capacity parameter, ξ_i represents parameters for handling the non-separable points, and index *i* shows the labels of N training sets [63].

If the dataset is just too hard more shown in Figure 5 then convert this dataset into higher dimensional space. The linear classifier relies on a dot product vector which is $K(X_i, X_j)$. Therefore, every single data point is mapped into higher dimensional space through a transformation process like $\phi: X \rightarrow \varphi(X)$ where a kernel function corresponds to an inner product within the expanded feature space. If data is not linearly separable, there is a function used for the transformation process to convert this data into higher dimensional space [64]. So, the data goes from 1D to 2D for representation of the data. Based on the function f(x), conversion of the dataset from 2D space to 2D feature dimensional space is easy. Now the only problem with transformation into higher dimensional feature space is that it's computationally expensive. Based on expensive calculations, researchers use a kernel trick to reduce the computational process. A function that takes as an input vector in the original space and returns the dot product of the vectors in the feature space is called a kernel function also referred to as kernel trick [65]. Using a kernel function, researchers can apply the dot product within the two vectors, so that every point is mapped to a higher dimensional space via some transformation. Therefore, essentially various scholars use this kernel trick to transform a non-linear space into a linear space. Some popular kernel tricks are used to transform the data into higher dimensional feature space [66].

- Linear kernel function, $K(X_i, X_j) = X_i^T . X_j$
- Polynomial kernel function, $K(X_i, X_j) = (1 + X_i^T . X_j)^p$
- Radial basis function (RBF), $K(X_i, X_j) = exp\left(-\frac{\|X_i - X_j\|^2}{2\sigma^2}\right)$
- Sigmoid kernel function, $K(X_i, X_j) = \tanh(\beta_0 X_i^T . X_j + \beta_1)$

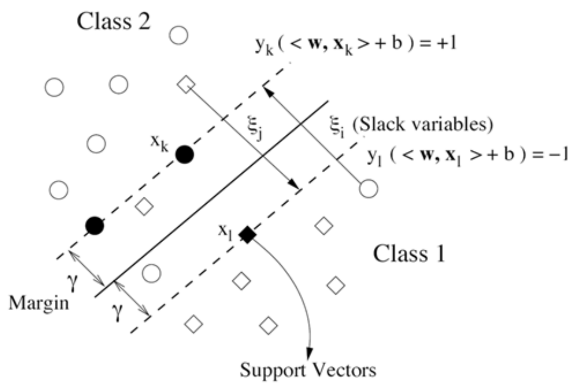

Figure 5. Linearly separable hyper-plane with slack variables

Unfortunately, choosing the correct kernel is a non-trivial task or maybe an unspecific task at hand, no matter which kernel researchers choose according to their problem. There are some properties used during the selection of kernel function e.g., the need to tune the kernel program to get good performance for the classifier. A famous program the researchers need to tune includes k-fold cross-validation [57]. Mercer's theorem: according to machine learning [67], kernel function scheme is a most famous trick that is a much near type of Mercer in which a variety of issues such as regression, classification, and inverse issues regarding optimization can be resolved competently. Kernel functions are linked along feature mapping due to its mapping procedure where the dataset is mapped from the original space to higher dimensional feature space. A common assumption where an input space X is mapped via a feature mapping $\{\Phi: X \rightarrow H | (x, y) \in X\}$, where $K(x, y) \langle \emptyset(x), \emptyset(y) \rangle H$. In Figure 6, this theorem performance steps described in detail with useful equations and matrices [46].

Input: Let us denote th_2 as the threshold value for R_{SVM} (p) selection in algorithm **step** (b) and th_2 as the threshold to select R_{SVM} (p) in algorithm step (c).
R_{SVM} (SVM $_{sent}$,p) **:** Set of SVM results obtained after performing SVM classification, SVM $_{sent}$ present the sentiments , p is the probability of sentence classification,
R_{NB} (NB_{sent} ,v) **:** Set of NB classification results obtained after performing NB classification,
NB_{sent} indicate sentiment: V-NB results value, contain "1'' for positive sentence and "-1'' for negative sentence.
Th3-min(R_{SVM} (p))

Figure 6. Stepwise Mercer's theorem

Where dual problem formulation for non-linear SVMs is described in (11) and (12) [68].

$$\{?\,(\alpha_i, \dots, \alpha_N) | max\big(Q(\alpha) = \sum \alpha_i - 1/2 \sum\sum \alpha_i \alpha_j y_i y_j k(X_i, X_j)\big), \sum \alpha_i y_i = 0 \ and\ \alpha_i \geq 0, \forall \alpha_i\} \quad (11)$$

$$f(x) = \sum \alpha_i y_i k(X_i, X_j) + b \quad (12)$$

4. RESULT

In this section, we present the different modified sections that took place in SVM by different research in different years. Table 2 (see in Appendix) presents the different categories of the result like the year, function, improvement, and objective of many researchers' modification in SVM for their research purpose [69]–[79]. In this research, we collect all the information related to SVM. Consequently, Table 3 shows some properties of SVM related to the applications, pros, and cons, and mentions in which section they are modified and for what purpose they are used all the data are mentioned in the given table where different research can be used.

Table 3. SVMs pros, cons, and applications

Parameter	Value
Pros: some advantages of SVMs are:	– They are effective in high dimensional spaces; they are still effective in cases where 'R' is the number of dimensions greater than the number of samples. According to the notations representation is: $\Rightarrow R^{\wedge}N > samples$. – They used a subset of training points in the decision function or support vectors. So, it's also memory efficient. – Support vectors are first it all. So different kernels can be specified for the decision functions. Common kernels are provided but it is also possible to specify the custom kernels. The addition of kernel functions together to achieve even more complex hyper-planes as $K1 + K2 = complex\ hyper - planes$, where K1 and K2 represents the first and second kernel functions respectively [79].
Cons: the disadvantages of SVMs include:	– SVM does not directly provide supervised probability estimates. – These are calculated using an expensive five-fold cross-validation. – If the number of features is greater than the number of samples then the method is likely to give poor performance.
Application: it can be a quite popular alternative to artificial neural networks. Some useful applications where SVMs used for good performance are:	– Medical imaging – Regression model to study the air quality in the urban areas – Image interpolation – Medical diagnosis task – Time series prediction as well as financial analysis – Encoding theory with practice – Pattern recognition – Page ranking algorithm – Text and object recognition

5. DISCUSSION

The optimization issues were resolved with the help of a support vector along with analytical analysis. Some conditions were applied to the selected dataset for further processing. Especially for small training datasets or the linear case, it is very critical to know which of the training datasets become support vectors. This situation can happen when the issue is relevant to the symmetry case. Generally, the worst-case computational complexity can happen during analytical analysis. Therefore, linear and non-linear SVM formulation provides the right path for the solution numerically and analytically. On the other hand, a range of approaches are used for the larger problems. Here in this article, we just describe the general research areas with the researcher's efforts and the generalization of the SVMs formulations for linear and non-linear problems. Some properties such as complexity, scalability, and parallelizability of SVMs play significant roles in the processing. The dataset of training and testing functions depends on the kernel functions although it corresponds to a dot product in higher dimensional space. There are some turning points for the SVMs where machine learning goes deep knowledge and requires more research such as the choice of the best kernel according to the dataset, processing speed, dataset size in terms of training and testing phases, rescaling, and optimal design for multi-class problems. According to the RBF kernel function, classifiers will automatically give values for the RBF weights, number of centers, center positions, and threshold.

6. CONCLUSION

This article provides a detailed description of the concept of linear and non-linear SVM for multiple areas of research such as sentiment analysis, text classification or categorization, image classification, and bio-informatics. It gives both hypothetical and mathematical verifications that SVMs show this is a very appropriate option for multiple fields. The hypothetical analysis concludes that SVMs accept the specific properties for completing the given tasks. The mathematical generalization shows the influence of SVMs on consistently achieving good performance, and advancements in existing methods considerably and drastically outperforming. In this paper, the authors also present the kernel trick idea of SVM in multiple fields with their four main functions linear, polynomial, RBF, and sigmoid functions. All under-reviewed articles show that the best kernel functions selection provides excellent results with high effects of accuracy. We analyze the field areas of the existing linear and non-linear SVMs with its kernel trick along selection of the function and parameters settings. We also analyze the concept that is related to the classical procedure of SVM for training the weights, improved training, and testing error rates through soft margin. The improvement in testing error or risk was exclusively the reason for the lower value of the training error or empirical risk. All this makes SVMs a very promising and easy-to-use method for learning multiple field classifiers from the given or selected examples.

APPENDIX

Table 2. Summary of result section with the help of improvements and objectives of the SVM functions

Ref.	Year	Function	Improvement	Objective
[69]	2000	RBF $K(x,y) = exp(-\gamma\|x-y\|^2)$ (4)	$|P(x-y)|^2 = (x-y)^T P.P(x-y)$1 $\|P(x-y)\|^2 = (x-y)^T S(x-y)$ 2 $f(x) = sgn\left(\sum_{i=1}^{i=l} \alpha_i y_i K(x,x_i) + b\right)$ 3 $K(x,y) = exp(-\gamma\|(x-y)^T S(x-y)\|^2)$ (5)	Semantic kernel function used for text categorization.
[70]	2015	Mahalanobis distance $M_{dist} = \sqrt{(x-m)C_x^{-1}(x-}$ (4)	$\phi(w) = \frac{1}{2}\|\vec{w}\|^2$ $subject\ to\ y_i(wx_i + b) \geq 1, \forall(x_i,y_i) \in D$ (1) $f(x) = sgn(\sum_{i=1}^{N\,s} \alpha_i y_i K(\vec{x}_i,\vec{x}) + b)$ 2 $K(\vec{x}_i,\vec{x}_i) = \phi(\vec{x}_i).\phi(\vec{x}_i)$ (3) $M_{dist} = \sqrt{(x-m)I^{-1}(x-m)^T} = \sqrt{(x-m)^2} = E_{dist}(x,m) = EDC$ (5) $d_i = \sqrt{\sum_{i=1}^{n}(x_i - m_i)^2}$ (6)	SVM was used in the training phase and M_{dist} was used in the testing phase for improving the classification accuracy.
[71]	2008	Reference in this paper 27: $\Psi(x) = \sum_{i=1}^{N} C_i\varphi_i(x)$ and $k(x,y) = \sum_{i=1}^{\infty} \lambda_i^2\, \varphi_i(x)\varphi_i(y)$	$wx + b = 0$ (4) $min\frac{1}{2}\|w\|^2 + C\sum_i \xi_i$ (5) With constraints $y_i(x_i w + b) \geq 1 - \xi_i,\ \xi_i \geq 0, \forall_i$ (6)	To investigate the effectiveness of using multi-words for text representation on the performances of text classification
[72]	2016	RBF $K(z_i,z_\tau) = exp\left(-\langle z_i - z_\tau, z_i - z_\tau\rangle / (2.\sigma^2)\right)$	$\langle w,w\rangle \to min, y_i.(\langle w,z_i\rangle + b) \geq 1, i = \overline{1,S}$ (1) $\left\{-L(\lambda) = -\sum_{i=1}^{S}\lambda_i + \frac{1}{2}.\sum_{i=1}^{S}\sum_{\tau=1}^{S}\lambda_i.\lambda_\tau.y_i.y_\tau.K(z_i,z_\tau) \to min_\lambda, \sum_{i=1}^{S}\lambda_i.y_i = 0, 0 \leq \lambda_i \leq C, i = \overline{1,S}\right\}$ (2) $f(z) = \sum_{i=1}^{S}\lambda_i.y_i.K(z_i,z) + b$ (3) $F(z) = sgn\left(f(z)\right) = sign(\sum_{i=1}^{S}\lambda_i.y_i.K(z_i,z) + b)$ (4)	SVM ensembles are used for big data classification with the modified particle swarm optimization.
[73]	2016		$\min_{w,b,\xi}\frac{1}{2}W^T W + C\sum_{i=1}^{l}\xi_i$ subject to $y_i(w^T \emptyset(X_i) + b) \geq 1 - \xi_i, \xi_i \geq 0$ (5)	To examine the accuracy of the SVM performance for the classification of human sentiments.

Table 2. Summary of result section with the help of improvements and objectives of the SVM functions

Ref.	Year	Function	Improvement	Objective
[74]	2015	RBF $K_{GAU}(x_i, x_j) = exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right), \sigma$ R_+ (3)	$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i , s.t. y_i(\langle w, \emptyset(x)\rangle + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall_i$ (1) $\max_{\alpha} \sum_{i=1}^{N} \alpha - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j \langle \emptyset(x_i), \emptyset(x_j)\rangle , s.t. \sum_{i=1}^{n} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, \forall_i$ (2) $\max_{\beta_i} \sum_{i=1}^{n} \beta_i K(x_i, x_i) - \sum_{i,j=1}^{n} \beta_i\beta_j K(x_i, x_j), s.t. \sum_{i=1}^{n} \beta_i = 1, \beta_i \geq 0, i = 1 \dots n$ (10) $R^2 = \beta^T diag(K) - \beta^T K\beta$ (11) $\frac{\partial T}{\partial \alpha_i^k} = R^2 * \frac{\partial\|w\|^2}{\partial \alpha_i^k} + \frac{\partial R^2}{\partial \alpha_i^k} * \|w\|^2$ (17) $\frac{\partial\|w\|^2}{\partial \alpha_i^k} = -\sum_{i,j=1}^{n} \alpha_i^* \alpha_j^* y_i y_j \frac{\partial K(x_i,x_j)}{\partial \alpha_i^k} = -\alpha^T\left(\frac{\partial K}{\partial \alpha_i^k}\right)\alpha$ (19) $\frac{\partial R^2}{\partial \alpha_i^k} = \sum_{i=1}^{n} \beta_i^* \frac{\partial K(x_i,x_i)}{\partial \alpha_i^k} - \sum_{i,j=1}^{n} \beta_i^* \beta_j^* \frac{\partial K(x_i,x_j)}{\partial \alpha_i^k} = \beta^T diag\left(\frac{\partial K}{\partial \alpha_i^k}\right) - \beta^T\left(\frac{\partial K}{\partial \alpha_i^k}\right)\beta$ (20) $\frac{\partial T}{\partial \alpha_i^k} = -R^2\left(\alpha^T \frac{\partial K}{\partial \alpha_i^k}\alpha\right) + \|w\|^2\left(\beta^T diag\left(\frac{\partial K}{\partial \alpha_i^k}\right) - \beta^T\left(\frac{\partial K}{\partial \alpha_i^k}\right)\beta\right)$(22) $\frac{\partial K(x,\acute{x})}{\partial \alpha_i^k} = exp\left(-\frac{\|x-\acute{x}\|^2}{2\sigma^2}\right)\left(-\frac{1}{2\sigma^2} \times \frac{\partial(\|x-\acute{x}\|^2)}{\partial \alpha_i^k}\right) = K(x, \acute{x})\left(-\frac{x^k-\acute{x}^k}{\sigma^2}\right) \times \begin{cases} +1\ if\ x^k = \alpha_i^k \\ -1\ if\ \acute{x}^k = \alpha_i^k \\ 0\ otherwise \end{cases}$ (24)	Numerical and nominal attributes are used in practical tasks, and the proposed heterogeneous SVM improves classification performance for both of them.
[75]	2016	RBF $k(x_i, x_j) = e^{-x_i - x_j^2/2\sigma^2}$ (2.6)	$Min \frac{1}{2} w^2\ s.t. y_i f(x_i) \geq 1 - \theta D(x_i)^2$ (2.8) $D(x_i)^2 = \frac{(x_i-\mu)^t S^{-1}(x_i-\mu)}{max(D(x_i)^2)}$ (2.9) $0 \leq D(x_i)^2 \leq 1$ (2.10) $y_i f(x_i) = 1 - \theta D(x_i)^2$ (2.11) $L_p = \frac{1}{2} w^2 - \sum \alpha_i (y_i f(x_i) - 1 + \theta D(x_i)^2)$ (2.12) $\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$ (2.13) $\frac{\partial L_p}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0$ (2.14) $w = \sum_{i=1}^{n} \alpha_i y_i x_i$ (2.15) $\alpha_i\{y_i f(x_i) - 1 + \theta D(x_i)^2\} = 0$ (2.16) $w(\alpha, \theta) = \sum_{i=1}^{l} \alpha_i (1 - \theta D(x_i)^2) - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i\alpha_j y_i y_j k(x_i, x_j)$ (2.17) $Max\ w(\alpha, \theta) = \sum_{i=1}^{l} \alpha_i (1 - \theta D(x_i)^2) - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i\alpha_j y_i y_j k(x_i, x_j)\ s.t. \sum_{i=1}^{l} \alpha_i y_i = 0, \alpha_i \geq 0$ (2.18) $\frac{\partial w(\alpha,\theta)}{\partial \theta} = -\sum_{i=1}^{l} \alpha_i \frac{(x_i-\mu_i)^t S^{-1}(x_i-\mu_i)}{max(D(x)^2)}$(2.19) $\hat{\theta}_{t+1} = \hat{\theta}_t + \rho\left(\sum_{i=1}^{l} \alpha_i \frac{(x_i-\mu_i)^t S^{-1}(x_i-\mu_i)}{max(D(x)^2)}\right), t \geq 0$ (2.20) $\frac{\partial w(\alpha,\theta)}{\partial \sigma^2} = -\left(\frac{1}{2}\sum_{i,j=1}^{l} \alpha_i\alpha_j y_i y_j x_i - x_j^2 k(x_i, x_j)\right)$ (2.21) $\hat{\sigma}_{t+1}^2 = \hat{\sigma}_t^2 + \rho\left(\frac{1}{2}\sum_{i,j=1}^{l} \alpha_i\alpha_j y_i y_j x_i - x_j^2 k(x_i, x_j)\right), t \geq 0$ (2.22)	To evaluate the effectiveness of the proposed modified slack variables within the SVM to solve complex data problems including class imbalance and overlapping.
[76]	2016	RBF $k(X_i, X_j) = exp\left\{\frac{\|X_i - X_j\|^2}{2\sigma^2}\right\}$ (9)	$\emptyset: (D_s, C_s) \to \{P, N\}$ (3) $\emptyset: R^N \to R^F$ (4) $\mathcal{L}(w, b, \alpha, \xi, \gamma) = \frac{\|w\|^2}{2} + c\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \|y_i(wx_i + b) - 1 + \xi_i\| - \sum_{i=1}^{n} \gamma_i \xi_i$ (5) $f(x) = sign\{\sum_{i=1}^{n} \alpha_i y_i (x_i \times x) + b\}$ (6) $\max\left[\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n} \alpha_i\alpha_j y_i y_j k(x_i . x_j)\right]$ (7) $f(x) = sign\{\sum_{i=1}^{n} \alpha_i y_i (x_i \times x) + b\}$ (10) where b is the solution of the equation $\alpha_i[y_i \sum_{i=1}^{n} \alpha_i y_i k(x_i \times x) + b - 1] = 0\ for\ non-zero\ \alpha_i$	A robust classification approach used for feature review's identification and semantic knowledge on the base of SVM and fuzzy domain ontology. The proposed approach increases. The precision rate of review's and opinion word's extraction and accuracy.
[77]	2013	Linear function	$R_t = +1\ if\ x_t * C1\ and$ $R_t = -1\ if\ x_t * C2$ $wt. x_t + w_0 \geq +1\ if\ x_t \in C1$ $wt. x_t + w_0 \geq -1\ if\ x_t \in C2$	Sentiment classification on online movie reviews through three approaches and out of these, SVM performs best in term of accuracy.
[78]	2014	RBF $k(X_i, X_j) = exp\left\{\frac{\|X_i - X_j\|^2}{2\sigma^2}\right\}$ (5)	$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i\ s.t. y_i.(w^T.x_i + b) \geq 1 - \xi_i, i = 1 \dots. m, \xi_i \geq 0, i = 1 \dots m$ (2) $f(x) = sign(\sum_{i=1}^{m} y_i \alpha_i^* K(x, x_i) + b^*)$ (3) $\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,s=1}^{m} \alpha_i\alpha_s y_i y_s K(x_i, x_s)\ s.t. \sum_{i=1}^{m} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1 \dots. m$ (4)	To address feature selection and class imbalance problems and give simultaneous solutions for both problems.

Table 2. Summary of result section with the help of improvements and objectives of the SVM functions

Ref.	Year	Function	Improvement	Objective
[79]	2016	SVM structure framework	$W^2(\alpha) = \sum_{i,s=1}^{m} \alpha_i \alpha_s y_i y_s K(x_i, x_s)$ (7) $\hat{y} = \underset{y \in Y}{argmax} \left(F(x, y)\right)$ $\underset{w,\xi}{min} \frac{\|w\|^2}{2} + C\xi \; s.t. \forall(\bar{y}_1, \dots, \bar{y}_N) \in Y^N: \frac{1}{N}\langle w, \sum_{i=1}^{N} [\Psi(x_i, y_i) - \Psi(x_i, \bar{y}_i)]\rangle \geq \frac{1}{N}\sum_{i=1}^{N} \Delta(y_i, \bar{y}_i) - \xi$	Two hierarchical text categorization approaches are used for text documents with byte n-gram-based document representation.

REFERENCES

[1] T. Joachims, “Training linear SVMs in linear time,” in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug. 2006, pp. 217–226, doi: 10.1145/1150402.1150429.

[2] H. Q. Minh, P. Niyogi, and Y. Yao, “Mercer’s theorem, feature maps, and smoothing,” in *Lecture Notes in Computer Science()*, Berlin: Springer, 2006, pp. 154–168, doi: 10.1007/11776420_14.

[3] K. Bennett and E. Bredensteiner, “Duality and geometry in SVM classifiers,” in *ICML ’00: Proceedings of the Seventeenth International Conference on Machine Learning*, 2000, pp. 57–64.

[4] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the fifth annual workshop on Computational learning theory*, Jul. 1992, pp. 144–152, doi: 10.1145/130385.130401.

[5] H. W. Guggenheimer, *Applicable geometry: global and local convexity*. Huntington: R.E. Krieger Pub. Co., 1977.

[6] J. Moody, “The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems,” in *Advances in Neural Information Processing Systems*, 1991, vol. 4.

[7] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[8] C. J. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.

[9] A. Beutelspacher and U. Rosenbaum, *Projective geometry: from foundations to applications*. Cambridge University Press, 1998.

[10] D. Decoste and B. Schölkopf, “Training invariant support vector machines,” *Machine Learning*, vol. 46, no. 1–3, pp. 161–190, 2002, doi: 10.1023/A:1012454411458.

[11] P. Pavlidis, I. Wapinski, and W. S. Noble, “Support vector machine classification on the web,” *Bioinformatics*, vol. 20, no. 4, pp. 586–587, Mar. 2004, doi: 10.1093/bioinformatics/btg461.

[12] P. Mitra, B. U. Shankar, and S. K. Pal, “Segmentation of multispectral remote sensing images using active support vector machines,” *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1067–1074, Jul. 2004, doi: 10.1016/j.patrec.2004.03.004.

[13] L. Bruzzone, M. Chi, and M. Marconcini, “A novel transductive SVM for semisupervised classification of remote-sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006, doi: 10.1109/TGRS.2006.877950.

[14] Y. Bazi and F. Melgani, “Semisupervised PSO-SVM regression for biophysical parameter estimation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1887–1895, Jun. 2007, doi: 10.1109/TGRS.2007.895845.

[15] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, “Customizing kernel functions for SVM-based hyperspectral image classification,” *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 622–629, Apr. 2008, doi: 10.1109/TIP.2008.918955.

[16] A. Smola and S. Vishwanathan, *Introduction to machine learning*. Cambridge: Cambridge University Press, 2008.

[17] R. Zhang and J. Ma, “An improved SVM method P-SVM for classification of remotely sensed data,” *International Journal of Remote Sensing*, vol. 29, no. 20, pp. 6029–6036, Oct. 2008, doi: 10.1080/01431160802220151.

[18] F. Bovolo, L. Bruzzone, and M. Marconcini, “A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008, doi: 10.1109/TGRS.2008.916643.

[19] N. Ghoggali and F. Melgani, “Genetic SVM approach to semisupervised multitemporal classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 212–216, Apr. 2008, doi: 10.1109/LGRS.2008.915600.

[20] W. Zhang, T. Yoshida, and X. Tang, “Text classification based on multi-word with support vector machine,” *Knowledge-Based Systems*, vol. 21, no. 8, pp. 879–886, Dec. 2008, doi: 10.1016/j.knosys.2008.03.044.

[21] A. Shmilovici, “Support vector machines,” in *Data Mining and Knowledge Discovery Handbook*, Boston, MA: Springer US, 2009, pp. 231–247.

[22] L. Bruzzone and C. Persello, “A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2142–2154, Jul. 2009, doi: 10.1109/TGRS.2008.2011983.

[23] C. A. C. Coello, G. B. Lamont, and D. A. V. Veldhuizen, *Evolutionary algorithms for solving multi-objective problems*. Boston, MA: Springer US, 2007.

[24] U. Maulik and D. Chakraborty, “A self-trained ensemble with semisupervised SVM: An application to pixel classification of remote sensing imagery,” *Pattern Recognition*, vol. 44, no. 3, pp. 615–623, Mar. 2011, doi: 10.1016/j.patcog.2010.09.021.

[25] Q. Liu, C. Chen, Y. Zhang, and Z. Hu, “Feature selection for support vector machines with RBF kernel,” *Artificial Intelligence Review*, vol. 36, no. 2, pp. 99–115, Aug. 2011, doi: 10.1007/s10462-011-9205-2.

[26] S. Patra and L. Bruzzone, “A cluster-assumption based batch mode active learning technique,” *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1042–1048, Jul. 2012, doi: 10.1016/j.patrec.2012.01.015.

[27] U. Maulik and D. Chakraborty, “A novel semisupervised SVM for pixel classification of remote sensing imagery,” *International Journal of Machine Learning and Cybernetics*, vol. 3, no. 3, pp. 247–258, Sep. 2012, doi: 10.1007/s13042-011-0059-3.

[28] H. Zhang, W. Shi, and K. Liu, “Fuzzy-topology-integrated support vector machine for remotely sensed image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 850–862, Mar. 2012, doi: 10.1109/TGRS.2011.2163518.

[29] N. Deng, Y. Tian, and C. Zhang, *Support vector machines: optimization based theory, algorithms, and extensions*. Chapman and Hall/CRC, 2012.

[30] B. Gaonkar and C. Davatzikos, "Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification," *NeuroImage*, vol. 78, pp. 270–283, Sep. 2013, doi: 10.1016/j.neuroimage.2013.03.066.
[31] R. Parveen, N. Shrivastava, and P. Tripathi, "Sentiment classification of movie reviews by supervised machine learning approaches using ensemble learning & voted algorithm," in *2nd International Conference on Data, Engineering and Applications (IDEA)*, Feb. 2020, pp. 1–6, doi: 10.1109/IDEA49133.2020.9170684.
[32] Y. Kodratoff, *Introduction to machine learning*, Elsevier, 1989.
[33] Z. Sun, C. Wang, D. Li, and J. Li, "Semisupervised classification for hyperspectral imagery with transductive multiple-kernel learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 11, pp. 1991–1995, Nov. 2014, doi: 10.1109/LGRS.2014.2316141.
[34] S. Maldonado, R. Weber, and F. Famili, "Feature selection for high-dimensional class-imbalanced data sets using support vector machines," *Information Sciences*, vol. 286, pp. 228–246, Dec. 2014, doi: 10.1016/j.ins.2014.07.015.
[35] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical learning with sparsity*, Chapman and Hall/CRC, 2015.
[36] L. Barghout, "Spatial-taxon information granules as used in iterative fuzzy-decision-making for image segmentation," in *Granular Computing and Decision-Making. Studies in Big Data*, 2015, pp. 285–318.
[37] C. D. A. Vanitha, D. Devaraj, and M. Venkatesulu, "Gene expression data classification using support vector machine and mutual information-based gene selection," *Procedia Computer Science*, vol. 47, pp. 13–21, 2015, doi: 10.1016/j.procs.2015.03.178.
[38] M. H. Bhavsar and D. A. Ganatra, "Support vector machine classification using mahalanobis distance function," *International Journal of Scientific and Engineering Research*, vol. 6, no. 1, pp. 618–626, Jan. 2015, doi: 10.14299/ijser.2015.01.015.
[39] S. Peng, Q. Hu, Y. Chen, and J. Dang, "Improved support vector machine algorithm for heterogeneous data," *Pattern Recognition*, vol. 48, no. 6, pp. 2072–2083, Jun. 2015, doi: 10.1016/j.patcog.2014.12.015.
[40] F. Ali, K.-S. Kwak, and Y.-G. Kim, "Opinion mining based on fuzzy domain ontology and support vector machine: a proposal to automate online review classification," *Applied Soft Computing*, vol. 47, pp. 235–250, Oct. 2016, doi: 10.1016/j.asoc.2016.06.003.
[41] Z. Yin and J. Hou, "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes," *Neurocomputing*, vol. 174, pp. 643–650, Jan. 2016, doi: 10.1016/j.neucom.2015.09.081.
[42] L. Demidova, E. Nikulchev, and Y. Sokolova, "Big data classification using the SVM classifiers with the modified particle swarm optimization and the SVM ensembles," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, 2016, doi: 10.14569/IJACSA.2016.070541.
[43] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using N-gram machine learning approach," *Expert Systems with Applications*, vol. 57, pp. 117–126, Sep. 2016, doi: 10.1016/j.eswa.2016.03.028.
[44] Y.-C. Chen and C.-T. Su, "Distance-based margin support vector machine for classification," *Applied Mathematics and Computation*, vol. 283, pp. 141–152, Jun. 2016, doi: 10.1016/j.amc.2016.02.024.
[45] J. Graovac, J. Kovacevic, and G. Pavlovic-Lazetic, "Hierarchical vs. flat N-gram-based text categorization: can we do better?," *Computer Science and Information Systems*, vol. 14, no. 1, pp. 103–121, 2017, doi: 10.2298/CSIS151017030G.
[46] T. Alam, R. Gupta, S. Qamar, and A. Ullah, "Recent applications of artificial intelligence for sustainable development in smart cities," in *Recent Innovations in Artificial Intelligence and Smart Applications. Studies in Computational Intelligence*, Cham: Springer, 2022, pp. 135–154, doi: 10.1007/978-3-031-14748-7_8.
[47] K. Korovkinas and P. Danėnas, "SVM and Naïve Bayes classification ensemble method for sentiment analysis," *Baltic Journal of Modern Computing*, vol. 5, no. 4, Dec. 2017, doi: 10.22364/bjmc.2017.5.4.06.
[48] F. S. Al-Anzi and D. AbuZeina, "Toward an enhanced Arabic text classification using cosine similarity and latent semantic indexing," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 2, pp. 189–195, Apr. 2017, doi: 10.1016/j.jksuci.2016.04.001.
[49] U. Maulik and D. Chakraborty, "Remote sensing image classification: a survey of support-vector-machine-based advanced techniques," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 1, pp. 33–52, Mar. 2017, doi: 10.1109/MGRS.2016.2641240.
[50] A. V. Phan, M. Le Nguyen, and L. T. Bui, "Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems," *Applied Intelligence*, vol. 46, no. 2, pp. 455–469, Mar. 2017, doi: 10.1007/s10489-016-0843-6.
[51] A. H. Haghiabi, H. M. Azamathulla, and A. Parsaie, "Prediction of head loss on cascade weir using ANN and SVM," *ISH Journal of Hydraulic Engineering*, vol. 23, no. 1, pp. 102–110, Jan. 2017, doi: 10.1080/09715010.2016.1241724.
[52] M. A. Beer, "Predicting enhancer activity and variant impact using gkm-SVM," *Human Mutation*, vol. 38, no. 9, pp. 1251–1258, Sep. 2017, doi: 10.1002/humu.23185.
[53] X. Niu, C. Yang, H. Wang, and Y. Wang, "Investigation of ANN and SVM based on limited samples for performance and emissions prediction of a CRDI-assisted marine diesel engine," *Applied Thermal Engineering*, vol. 111, pp. 1353–1364, Jan. 2017, doi: 10.1016/j.applthermaleng.2016.10.042.
[54] P. Kumar and A. Tiwari, Eds., *Ubiquitous machine learning and its applications*, IGI Global, 2017.
[55] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Computer Science*, vol. 127, pp. 511–520, 2018, doi: 10.1016/j.procs.2018.01.150.
[56] S. Xie, Z. Li, and H. Hu, "Protein secondary structure prediction based on the fuzzy support vector machine with the hyperplane optimization," *Gene*, vol. 642, pp. 74–83, Feb. 2018, doi: 10.1016/j.gene.2017.11.005.
[57] D. Sebai and A. U. Shah, "Semantic-oriented learning-based image compression by only-train-once quantized autoencoders," *Signal, Image and Video Processing*, vol. 17, no. 1, pp. 285–293, Feb. 2023, doi: 10.1007/s11760-022-02231-1.
[58] A. Ullah, A. Salam, H. El Raoui, D. Sebai, and M. Rafie, "Towards more accurate iris recognition system by using hybrid approach for feature extraction along with classifier," *International Journal of Reconfigurable and Embedded Systems (IJRES)*, vol. 11, no. 1, pp. 59–70, Mar. 2022, doi: 10.11591/ijres.v11.i1.pp59-70.
[59] A. Ullah and A. Chakir, "Improvement for tasks allocation system in VM for cloud datacenter using modified bat algorithm," *Multimedia Tools and Applications*, vol. 81, no. 20, pp. 29443–29457, Aug. 2022, doi: 10.1007/s11042-022-12904-1.
[60] I. Guyon, V. Vapnik, B. Boser, L. Bottou, and S. A. Solla, "Structural risk minimization for character recognition," in *Advances in Neural Information Processing Systems*, 1991, vol. 4.
[61] D. Ciucci *et al.*, Eds., *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, vol. 1601. Cham: Springer International Publishing, 2022, doi: 10.1007/978-3-031-08971-8.
[62] R. Cuingnet *et al.*, "Spatial regularization of SVM for the detection of diffusion alterations associated with stroke outcome," *Medical Image Analysis*, vol. 15, no. 5, pp. 729–737, Oct. 2011, doi: 10.1016/j.media.2011.05.007.
[63] S. N. Khan *et al.*, "Comparative analysis for heart disease prediction," *JOIV : International Journal on Informatics Visualization*, vol. 1, no. 4–2, p. 227, Nov. 2017, doi: 10.30630/joiv.1.4-2.66.
[64] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their*

Applications, vol. 13, no. 4, pp. 18–28, Jul. 1998, doi: 10.1109/5254.708428.
[65] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, "SVM active learning approach for image classification using spatial information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 4, pp. 2217–2233, Apr. 2014, doi: 10.1109/TGRS.2013.2258676.
[66] U. Ullah and A. Ullah, "An evolutionary algorithm for the solution of multi-objective optimization problem," *International Journal of Advances in Applied Sciences*, vol. 11, no. 4, pp. 287–295, Dec. 2022, doi: 10.11591/ijaas.v11.i4.pp287-295.
[67] S. N. Khan, N. M. Nawi, M. Imrona, A. Shahzad, A. Ullah, and A.- Rahman, "Opinion mining summarization and automation process: a survey," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 5, p. 1836, Oct. 2018, doi: 10.18517/ijaseit.8.5.5002.
[68] A. Hanane, A. Ullah, and S. Raghay, "Enhanced GAF protocol based on graph theory to optimize energy efficiency and lifetime in WSN technology," *International Journal of Intelligent Unmanned Systems*, vol. 11, no. 2, pp. 214–225, Apr. 2023, doi: 10.1108/IJIUS-08-2021-0096.
[69] A. Ullah, I. Laassar, C. B. Şahin, O. B. Dinle, and H. Aznaoui, "Cloud and internet-of-things secure integration along with security concerns," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 12, no. 1, pp. 62–71, Apr. 2023, doi: 10.11591/ijict.v12i1.pp62-71.
[70] B. Saeeda, T. Noor, A. Ullah, and H. Aznaoui, "A critical study of the teaching attitudes of teachers at primary school level in Pakistan," *International Journal of Knowledge and Learning*, vol. 14, no. 4, p. 385, 2021, doi: 10.1504/IJKL.2021.118556.
[71] A. Ullah, S. Baseer, and S. Umar, "Role of cooperation in energy minimization in visual sensor network," in *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, Aug. 2016, pp. 447–452, doi: 10.1109/INTECH.2016.7845026.
[72] A. Ullah, N. M. Nawi, A. Arifianto, I. Ahmed, M. Aamir, and S. N. Khan, "Real-time wheat classification system for selective herbicides using broad wheat estimation in deep neural network," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no. 1, p. 153, Jan. 2019, doi: 10.18517/ijaseit.9.1.5031.
[73] N. Tenali and G. R. M. Babu, "A systematic literature review and future perspectives for handling big data analytics in COVID-19 diagnosis," *New Generation Computing*, vol. 41, no. 2, pp. 243–280, Jun. 2023, doi: 10.1007/s00354-023-00211-8.
[74] Q. I. Ahmed, H. Attar, A. Amer, M. A. Deif, and A. A. A. Solyman, "Development of a hybrid support vector machine with grey wolf optimization algorithm for detection of the solar power plants anomalies," *Systems*, vol. 11, no. 5, p. 237, May 2023, doi: 10.3390/systems11050237.
[75] T. Saida and M. Nishio, "Transfer learning Gaussian process regression surrogate model with explainability for structural reliability analysis under variation in uncertainties," *Computers & Structures*, vol. 281, p. 107014, Jun. 2023, doi: 10.1016/j.compstruc.2023.107014.
[76] J. Niyogisubizo *et al.*, "Predicting traffic crash severity using hybrid of balanced bagging classification and light gradient boosting machine," *Intelligent Data Analysis*, vol. 27, no. 1, pp. 79–101, Jan. 2023, doi: 10.3233/IDA-216398.
[77] S. A. and S. R., "A systematic review of explainable artificial intelligence models and applications: recent developments and future trends," *Decision Analytics Journal*, vol. 7, p. 100230, Jun. 2023, doi: 10.1016/j.dajour.2023.100230.
[78] A. H. Elsheikh, "Applications of machine learning in friction stir welding: prediction of joint properties, real-time control and tool failure diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 121, p. 105961, May 2023, doi: 10.1016/j.engappai.2023.105961.
[79] A. Ullah, S. N. Khan, and N. M. Nawi, "Review on sentiment analysis for text classification techniques from 2010 to 2021," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 8137–8193, Mar. 2023, doi: 10.1007/s11042-022-14112-3.

BIOGRAPHIES OF AUTHORS

Sundas Naqeeb Khan is a researcher and she has worked as a lecturer, instructor, and adjunct professor with companies, academies, and universities. She has reviewed six double-blind peer reviews. She has worked as an invited researcher with national and international universities. Her broad research interests cover topics relating to electrical and electronics, computer and software engineering, optimization, data science, communication, and physics. She can be contacted at email: sndskhan87@gmail.com.

Samra Urooj Khan is working as a lecturer in Department of Electrical Engineering Technology, Punjab University of Technology, Rasul, Mandi Bahauddin, Punjab, Pakistan. Her academic qualification is M.S. in Electrical Engineering from UTHM, B.S. from Pakistan. Her research area includes wireless communications, image processing, and signal processing. She can be contacted at email: samrauroojkhan1615@gmail.com.

Hanane Aznaoui works at Faculty of Computer Science, Cadi Ayyad University, Morocco. Her main research interests focus on bio-inspired computing, artificial intelligence, metaheuristic modeling and optimization, evolutionary computations, optimization algorithms, information retrieval, feature selection, combinatorial problems, data mining, and text mining. She can be contacted at email: h.aznaoui@gmail.com.

Canan Batur Şahin works at as assistant professor at Faculty of Computer Science, Turgut Özal University Turkey. Her research interests include software engineering, artificial intelligence, and optimization. She can be contacted at email: canan.batur@ozal.edu.tr.

Özlem Batur Dinler works as assistant professor at Faculty of Computer Science, Siirt Üniversitesi Turkey. Her research interests include artificial intelligent, machine learning, and deep learning and software engineering. She can be contacted at email: o.b.dinler@siirt.edu.tr.