

Article

# Complex and Hypercomplex-Valued Support Vector Machines: A Survey

Nancy Arana-Daniel 

CUCEI, University of Guadalajara, Guadalajara Jalisco C.P. 45400, Mexico; nancyaranad@gmail.com

Received: 5 June 2019; Accepted: 29 July 2019; Published: 31 July 2019



**Abstract:** In recent years, the field of complex, hypercomplex-valued and geometric Support Vector Machines (SVM) has undergone immense progress due to the compatibility of complex and hypercomplex number representations with analytic signals, as well as the power of description that geometric entities provide to object descriptors. Thus, several interesting applications can be developed using these types of data and algorithms, such as signal processing, pattern recognition, classification of electromagnetic signals, light, sonic/ultrasonic and quantum waves, chaos in the complex domain, phase and phase-sensitive signal processing and nonlinear filtering, frequency, time-frequency and spatiotemporal domain processing, quantum computation, robotics, control, time series prediction, and visual servoing, among others. This paper presents and discusses the importance, recent progress, prospective applications, and future directions of complex, hypercomplex-valued and geometric Support Vector Machines.

**Keywords:** complex; hypercomplex; Support Vector Machines

## 1. Introduction

The base idea of Support Vector Machines is the soft margin classifier that was first introduced by Cortes and Vapnik in 1995 [1], then in the same year, this algorithm was extended for deal with regression problems by Vapnik [2]. The first formal statistical analysis about boundaries of the generalization of hard margin SVMs were given by Barlett [3] and Shawe-Taylor, et al. in 1998 [4]. In 2000, Shawe-Taylor and Cristianini continued with the development of the statistical learning theory that supports the SVMs and they present the statistical bounds of the generalization of soft margin algorithms for the regression case [5]. Thus, since its introduction by Vladimir Vapnik and his team, SVMs have become an important and very used learning algorithm to solve regression and classification problems. The sheer number of published applications of support vector machines (SVM) for solving problems involving real data is reflected in the high number of citations that the top 100 SVM publications [6] have recorded; there are more than 46,000 citations, according to Google Scholar.

SVMs use an intelligent and elegant technique to solve non-linear problems: the kernel trick, which has become a seminal idea to develop several machine learning methods called kernel-based methods. Kernels are symmetric functions that for two vectors of dimension  $n$ , return a real number  $\neq 0$ . The use of kernels in SVM allows implicitly mapping the original input data into a higher dimensional Reproducing Kernel Hilbert Space (RKHS)  $H$ , in which linear functions that split the mapped data are computed. This is equivalent to solving the non-linear problem in the original input data space.

Although the RKHS includes complex and hypercomplex spaces in which the original input data can be mapped via kernel functions, the majority of extensions and applications of SVM deal with real data, and therefore use real kernel functions. However, the need for processing complex and hypercomplex data in modern applications has paved the way for some important proposals to

process complex and hypercomplex-valued input and output data, and kernels. These approaches will be reviewed in more detail in the next section.

Therefore, since the first proposals of SVM extensions for processing complex and hypercomplex data and kernels, their applications have grown in number, including signal processing, pattern recognition, classification of electromagnetic signals, light, sonic/ultrasonic and quantum waves, chaos in the complex domain, phase and phase-sensitive signal processing and nonlinear filtering, frequency, time-frequency and spatiotemporal domain processing, quantum computation, robotics, control, pattern recognition, time-series prediction, and visual servoing, among others. Due to the variety and importance of the applications of this type of SVM extension, in this survey we present and discuss the importance, recent progress, prospective applications, and future directions of complex, hypercomplex-valued and geometric support vector machines.

This survey is organized as follows: In Section 2, the most important extensions of SVM are reviewed to deal with complex and hypercomplex-valued input and output data, as well as with kernels that process this type of data. In Section 3, some of the applications of these extensions of SVMs with the greatest impact are introduced, as well as some of their promising developments with future applicability. The last section is devoted to the conclusions.

## 2. Complex and Hypercomplex Support Vector Machines Extensions

In this section, the most important extensions of SVM are reviewed to deal with complex and hypercomplex-valued input and output data, as well as with kernels that process this type of data. Each approach is studied in the chronological order of publication. In addition, Table 1 summarizes the information presented in all the subsections of this section, which presents the main features, advantages and disadvantages of each approach.

Table 1. Part 1-Complex and hypercomplex SVM extensions.

Ref./Year Section	Theoretical Frameworks Used	Advantages of the SVM Extension	Disadvantages of the SVM Extension
[7]/2000, [8]/2001 Section 2.1	Geometric/Clifford Algebra and quadratic programming	MIMO classification/regression can be performed  They consider input and output multivector-valued data (complex and quaternionic)	They did not consider the multivector-valued kernel design, or define the Gram matrix (or kernel matrix) as a multivector matrix.
[9]/2010, Section 2.1	Geometric/Clifford Algebra and quadratic programming	MIMO classification/regression can be performed and the approach can deal with dynamic systems or recurrence  It can be extended to complex, quaternions, octonions, dual quaternions, super-quadratic algebras and any GA of any dimension using the same methodology that it is shown in the paper The GA framework preserves the topology and geometry of the complex and hypercomplex input data and it is a coordinate-free framework	The GA framework it is not popular and it has to be taught to propose other GA extensions and to design GA kernels

Table 1. Cont.

Ref./Year Section	Theoretical Frameworks Used	Advantages of the SVM Extension	Disadvantages of the SVM Extension
[10]/2002, [11]/2004, Section 2.2.1	Division algebras (complex) and IRWLS	<p>MIMO regression can be performed</p> <p>It is applied to frequency nonselective channel estimation</p> <p>It deals with complex signals The optimization problem defined uses an <math>\epsilon</math>-insensitive cost function that in tur uses the <math>L - 2</math> norm to consider to consider all the multiple dimensions in a unique restriction</p>	<p>The output data are represented as real vectors.</p> <p>The optimization problem may become an ill-defined one, but it can also be solved using IRWLS</p>
[12]/2007, Section 2.2.2	Division algebras (real, complex and quaternions) and bilevel quadratic optimization	<p>The use of a cost function rotationally invariant was proposed in order to take into account the magnitude and not the angle of the error in the output.</p> <p>It presents real, complex and quaternionic-valued derivation of a well-defined dual optimization problem. The outputs of the SVR were interpreted as complex/quaternion-valued vectors to tackle the interconnectedness of outputs.</p>	<p>The experimental results are limited to show an equalization of a 4-QAM signal over a complex linear communication channel and complex function approximation</p>
[13]/2012, Section 2.2.4	Division algebras (octonions)	<p>It deals with octonion-valued regression</p> <p>The output and input data is expressed as pure octonions</p> <p>The proposal was applied to solve the gait analysis.</p>	<p>The derivation of the primal and dual optimization problem must consider the nonassociativity of the octonions and this adds complexity to the derivation of the dual problem, the training machine form and the computation of the kernel matrix</p> <p>As the nonassociativity feature adds terms to the loss function, then this increase the dimension of the problem and the curse of dimensionality could lead to computationally intractable problems for real applications</p>

Table 1. Cont.

Ref./Year Section	Theoretical Frameworks Used	Advantages of the SVM Extension	Disadvantages of the SVM Extension
[14]/2015, Section 2.2.5	Wirtinger’s calculus and widely linear mean square estimation	The Wirtinger calculus includes definitions of complex derivatives and they are used to compute the dual optimization problem of the SVM in an elegant and compact manner. The widely linear mean square estimation is a methodology to deal with complex numbers and when it is used with this kind of data it was proved that it can yield significant improvements in the estimation performance. The proposal was applied to solve problems of function estimation for a sinc function, channel identification, channel equalization and classification.	It is limited to complex data  For classification problems (MNIST classification) the error rate increased with the use of this proposal regarding using one-versus-three and the one-versus-one real SVM strategies.
[15]/2016, Section 2.1	Quaternion algebra and quadratic programming	It can be performed MIMO classification.  Two quaternion-valued kernels and a quaternion-valued <i>sign()</i> decision function were designed.	It is an approach restricted to be applied using quaternion-valued input and output data
[15]/2016, Subsec. 2.1	GA and quadratic programming	It can be performed MIMO classification.  A methodology to parallelize a CSVM using Gaussian kernel was presented to decrease the execution time of the CSVM.	It is an exclusive parallelization methodology for CSVM that uses a Gaussian kernel.

2.1. Geometric or Clifford Support Vector Machines

In 2000, the first attempt to develop an extension of SVM to deal with complex and hypercomplex input and output data was published [7,8]. The approach was designed using the Geometric Algebra (GA) framework [16].

GAs are the Clifford special family of associative algebras built over a field, and a quadratic space of elements called multivectors that is constructed using a special product called the geometric product. In Clifford and GAs we can find embedded concepts of complex, linear, tensor, quaternion, and octonion algebras (for a detailed introduction to Clifford and GAs it is recommended to read [16,17]).

Hence, in [7,8], a multivector-valued input and output data was used to perform multiclassification and multi-regression, solving one multivector-valued optimization problem. Nevertheless, the approach was not completely developed, because it did not consider the multivector-valued kernel design, or define the Gramm matrix (or kernel matrix) as a multivector matrix. Therefore, the design can be considered to be an application of the idea of Eiichi Goto presented in 1954 with the proposal of “Parametron” [18,19] wherein the phase of a high-frequency carrier is used to represent binary or multivalued information. Similarly, in [7,8], the elements of different grades of the multivector are used to represent multivalued information in the input and the output

multivector-valued data. It was not until 2010 that the extension presented in [7,8] was fully developed, with the introduction of Clifford Support Vector Machines (CSVMs) [9].

The author considers this to be the most complete work that generalizes the real-valued SVM into a complex and hypercomplex-valued SVM, as the Clifford Algebra theory includes concepts of the most used and important algebras, such as analytic geometry of Descartes, the complex algebra of Wessel and Gauss, Hamilton algebra, matrix Cayley algebra, exterior Grassmann algebra, tensor algebra of Ricci, the algebra of Pauli and Dirac, Lie algebra, etc. Furthermore, all the division algebras such as real, complex, quaternion and octonion can be viewed as isomorphic Clifford algebra.

CSVM presented the entire design of a multivector-optimization problem, the derivation of the primal and dual multivector-valued problems and the definition of the Gramm matrix as a matrix with multivectors as elements, as well as the introduction of a multivector-valued kernel design. The key feature of CSVM design is the use of the Clifford product involved in the multivector-valued kernels to keep separated the different grade components of each input multivector and at the end, to represent those components as the direct sum of linear spaces in multivector-valued outputs. All the above allow the CSVM to solve optimization problems in complex and hypercomplex vector spaces, so that the original and the feature (or RKHS) spaces can preserve the topology and geometry of the complex and hypercomplex input data. It is demonstrated that when this machine learning algorithm is used to process this type of data, the accuracy and convergence speed can be improved with respect to the obtained accuracy and convergence speed with real SVMs.

The CSVM was presented to solve multi-classification and multi-regression, and using it in conjunction with a long short-term memory neural network [20] can deal with dynamic systems or recurrence.

Two multivector-valued kernels are designed: a complex kernel using the GA  $G_{0,1,0}$  and a quaternion-valued kernel embedded in the GA  $G_{0,2,0}$ . The design methodology can be extended to any Clifford algebra (so it can be applied to solve any hypercomplex SVM problem).

This approach was used to solve multi-classification for object recognition, multi-interpolation, time-series forecasting, and path-planning problems.

In [15] a Quaternion Support Vector Machine (QSVM) was presented as an special case of the CSVM of [9]. The quaternion algebra framework was used to design a QSVM that processes multiple multivectors as input and returns multiple multivectors as output data to achieve multi-class classification. Two quaternion kernels that involve the quaternion product in their definition were designed: polynomial quaternion-valued kernel and Gaussian quaternion Gabor kernel. In addition quaternion  $sign()$  function is defined to allow the QSVM to classify up to sixteen classes using one QSVM. Diamond colour classification and object recognition experiments were conducted to demonstrate the algorithm efficiency.

The article [21] the methodology to implement a parallel CSVM using the Gaussian kernel was presented. As the Gaussian kernel returns a real number even when it is used with multivectors as input data, and in addition it has the commutative property (as the Gaussian kernel does not involves the Clifford product computation) the multivector input data can be separated into its different grade elements that belong to independent subspaces and therefore a quadratic optimization problem for each element can be solved using parallelism. Experiments of classification were conducted using benchmark problems such as the concentric 2D spirals and a five-class problem described with three overlapped circles.

## 2.2. Division Algebras

Several efforts using division algebras were made to generalize real SVMs, to deal with complex and hypercomplex numbers. Division algebras are appealing mathematical frameworks to solve the complex, hypercomplex-valued SVM optimization problem due to the fact that all non-zero elements of a vector space have multiplicative inverses. There are four division algebras: real numbers, complex numbers, quaternions, and octonions.

### 2.2.1. SVM Multiregression for Nonlinear Channel Estimation in Multiple-Input Multiple-Output Systems

In 2002, an approach to solve multidimensional function approximation and regression was presented [10], and was fully developed and published in 2004 [11]. The authors addressed the problem of multiple-input, multiple-output (MIMO) system for the frequency nonselective channel estimation. This was an SVM-based approach as it leveraged the multidimensionality of a MIMO channel by using a regression SVM-based algorithm. This study proposed using an iterative reweighted least square (IRWLS) to solve the multivariable SVM problem instead of quadratic programming, and train an SVM multi-regressor to model nonlinearities that affect each transmitter or each receiver module of the transmission-reception chain. These cases were modeled separately using two different channel models of the input information signal represented by the quadrature phase shift keying (QPSK) employed in complex signal processing. On the other hand, the output data are represented as real vectors and the optimization problem defined uses an  $\epsilon$ -insensitive cost function that in turn uses the  $L_2$  norm to consider all the multiple dimensions in a unique restriction. When  $\epsilon = 0$  the problem is equivalent to solving one independent, regularized real kernel for each dimension, but for  $\epsilon \neq 0$ , the problem becomes an ill-defined one that cannot be solved straightforwardly, and it takes an iterative procedure to obtain the solution which is why they use an IRWLS algorithm.

Simulations were conducted to address the problem of nonlinear channel estimation, and their proposal outperforms the applications of one real support vector regression (SVR) and a radial function network (RBFN) for each dimension. Meanwhile, for linear channel models, results equivalent to those obtained using minimum mean square error (MMSE) strategy were achieved.

### 2.2.2. Quaternionic and Complex-Valued Support Vector Regression for Equalization and Function Approximation

In [12], another SVR strategy was designed by Shilton et al., this time to deal with quaternionic and complex-valued equalization and function approximation. The outputs of the SVR were interpreted as complex/quaternion-valued vectors to tackle the interconnectedness of outputs, i.e., the fact that these outputs can be coupled and treated independently can lead to a decrease in regressor accuracy. The author proposed the use of a rotationally invariant cost function in order to consider the magnitude, but not the angle of the error in the output. When a real regressor is applied on each dimension of a complex or a quaternionic-valued signal, each one of the regressor estimates one function in its own axis, and when these functions are summed to construct a complex or quaternion output, the overall risk function will not be rotationally symmetric because it will contain the magnitude and the various angles added (even when the angles were computed using different axis for each regressor). Therefore, considering only the magnitude of the error but not the various angles between the axes, will affect the accuracy of the estimated function. This is one of the main reasons why complex and hypercomplex-valued estimators increase the accuracy when dealing with complex and hypercomplex-valued input data regarding the strategies that cast these problems on to the real domain by splitting data in each different grade part (two for complex, four for quaternion and so on), and then work with each dimension independently instead of solving the optimization problem in complex and hypercomplex input, output, and feature spaces. The authors of [12] derive a well-defined dual optimization problem using a bilevel definition of the primal convex quadratic problem [22]. In the experimental results section, they consider the problem of equalization of a 4-symbol quadrature amplitude modulated (4-QAM) signal over a complex linear communication channel added with Gaussian noise. Again, the comparison results show that their proposal outperforms two independent SVR with  $\epsilon$ -insensitive loss function and the decision boundary obtained with their approach is compared “favorably” with the optimal decision boundary that a Bayesian equalizer gets.

### 2.2.3. A Division Algebraic Framework for Multidimensional Support Vector Regression

In 2010 Shilton et al. tackled the multidimensional SVR design using the division algebras framework, to present an  $\epsilon$ -insensitive loss function independent of the coordinate system or basis

used [23]. Although the octonions are included in the division algebras, this paper only deals with real, complex and quaternions extensions of SVR. Indeed, the authors only present the design of a quaternionic  $\epsilon_H$ -SVR because the real and complex extensions are cases included as restrictions of the quaternionic SVR. The approach proposes the use of a strictly L2-norm-based loss function to avoid that the trained machine could be influenced by the axis choices, due to the fact that this loss function is coordinate independent and less sensitive to outliers than a quadratic cost function. Additionally, the L2 norm includes an  $\epsilon$ -insensitive region that ensures sparseness. The authors present the derivation of a primal quaternionic SVR using this loss function, and then the Lagrange multipliers optimization method is used to compute the quadratic-valued dual training problem, which at the end, is analogous to the standard real SVR dual, but considering the use of a quaternionic-valued kernel function. Their proposal is applied to solve problems of complex-valued function approximation, chaotic time-series forecasting and channel equalization. Experimental results show that the proposal outperforms Clifford SVR (C-SVR) presented in 2000 and 2001 [7,8], least-square SVR (LS-SVR) [24] and multidimensional SVR (M-SVR) [11].

Furthermore, a useful comparative analysis against C-SVR [7,8], LS-SVR [24], M-SVR [11], extreme learning machine (ELM) [25,26], and a kernel method [27] is performed in terms of the definition of primal loss functions, dual optimization problem, independence of coordinates of the loss functions, sparsity of solution, and outlier sensitivity. It is important to note that the methodologies of the comparative analysis were proposed in 2000, 2002, 2004, 2006, and 2002 respectively.

#### 2.2.4. A Note on Octonionic Support Vector Regression

Mentioned in the above section, Shilton et al., in their 2010 paper [23], did not include the design of an octonion-valued SVR. Instead, they include this design in a publication in 2012, [13]. Thus, in this extension to octonionic-valued SVR, to derive the primal and dual optimization problems the authors use the same methodology as the one used in [23]. The dual training problem obtained is similar to that presented for quaternionic-valued SVR; however, in the case of octonions, their nonassociativity feature had to be considered to this derivation. Therefore, the dual-training problem considers an antiassociator operator affecting the Kernel matrix term of the dual problem, as well as a subtracting quadratic term involving the Lagrange multipliers and an “associated kernel function” that computes a pondered  $\epsilon$ -summation of the kernel results and returns a real number. The same happens with the training machine form between the quaternion and octonion-valued SVR, for this last one, a term that it is included as a multiplier of the value of the associated kernel function is involved. Therefore, although their proposals for derivations of quaternionic and octonionic-valued SVRs are similar, the important difference in antiassociativity of octonions as regards the quaternions has to be factored in to compute the dual training problem, as well as the forms of training machine. This differs from the cases of the CSVM Section 2.1 and from the proposal presented in the next subsection, in which the generic and invariant methodologies of derivation of the dual optimization problem are shown independent of the algebra with which one works.

The paper [13] also shows three special cases of octonionic SVR in which the nonassociative octonion-valued loss-function-related terms can be neglected to obtain a “full-analogy” (not identity) between their proposals of real, complex, quaternion, and octonion SVR. These special cases were called pseudoreal, pseudocomplex, and pseudoquaternion regressions. The authors analyzed the features of a nonlinear function that maps data from the input space to the feature space (feature map or kernel function), to describe the cases when the associated kernel function terms are equal to zero, and the full analogy between their octonion-valued and quaternion-valued SVR proposal is achieved. Nevertheless, in the general octonion-valued SVR case, the curse of dimensionality could lead to computationally intractable problems depending on the number of input data, and therefore, this could render it impractical in real applications. However, this was not the case in the experimental results section.

Experimental results demonstrate the performance of the octonionic SVR when applied to the biomechanical study of human locomotion, i.e., gait analysis. In [28] a high correlation between inertial measurement systems (IMUs: accelerometers and gyroscopes) during human locomotion is suggested. However, due to the fact that the IMUs suffer from intrinsic drift error, there is a need to correct the sensor measurements or systems calibration. One way to achieve this is to use an intelligent regression to estimate the correct measurements of key events of the endpoint foot trajectory during human treadmill walking. This was performed by octonionic SVR, C-SVR, LS-SVR, and M-SVR. The targets of regression were expressed as pure octonions and, as the data sets contain a significant number of outliers, the octonionic SVR and the C-SVR obtained the best overall error, and the first one shows more consistency in obtaining better results.

### 2.2.5. Complex Support Vector Machines for Regression and Quaternary Classification

In [14] the derivation and design of a complex SVM for regression and classification are presented. The authors use two main frameworks to develop their proposal: the widely linear mean square estimation and Wirtinger's calculus.

In [29] a mean square estimation (MSE) optimization methodology to deal with complex and normal data is presented. The MSE for complex data is not linear as in the real case, but wide sense linear: i.e., the regression of a scalar random variable  $y$  to be estimated is linear, both in terms of a random vector  $x$  that is an observation of  $y$  and regarding the complex conjugate of  $x$ , namely  $x^*$ . This optimization method shows advantages regarding the linear procedure when is used with complex data, and it can yield significant improvements in the estimation performance as it was shown in the examples presented in [29].

Wirtinger's calculus was introduced in 1927 [30]. The Wirtinger derivatives (or Wirtinger operators) are first-order partial differential operators defined by several complex variables. These derivatives behave analogously to the ordinary derivatives of functions of real variables.

Thus, in the paper [14], the above frameworks were used in order to work with the SVMs in the complex Reproducing Kernel Hilbert Space (RKHS). Three approaches were presented:

(1) The *complexification* of the Hilbert space  $H^2 = H \times H$ : the objective is to enrich  $H^2$  with a complex inner product that is equivalent to the definition of the Clifford (or geometric) product of two complex numbers in a Clifford algebra isomorphic to the complex space. Hence, this approach is equivalent to work with this type of hypercomplex number in Clifford or GA algebras of Section 2.1.

(2) The Dual Real Channel (DRC) approach in which the training data are split into two sets (real and imaginary parts of the input data), and a real regression (or classification) is performed on each data set using a real kernel. It has been proven before that this approach and the complexification are equivalent procedures [31]

(3) The pure complex SVM, in this paper is derived using Wirtinger's calculus and widely linear estimation in a very elegant and compact manner. This derivation allows the authors to conclude that the proposed pure complex SVM can be solved by splitting the labels (desired outputs) of data into their real and imaginary parts, and solving two real SVM tasks employing any one of the standard algorithms (they employed SMO in their experiments). Hence, the difference between practical implementations of complexification, DRC and pure complex SVMs is only that the first two mentioned approaches use real kernels meanwhile the pure complex SVM approach uses an induced real kernel  $\kappa_C^r$ .

One of the main contributions of the paper is the validation of complex kernels to classify the target space into four different categories. This is a very intuitive result, as for real valued kernels the classification features two classes, and the complex kernel features a real and an imaginary part  $K = Kr + jKi$ , so there are four possible combinations for the signs of  $Kr$  and  $Ki$ , hence yielding four categories naturally.

The experiments were conducted using pure complex SVR and a complex-induced Gaussian kernel (which is not equal to the real Gaussian RBF) for the function estimation of a sinc function,



channel identification, and channel equalization. In addition, a multiclass classification application was demonstrated using the MNIST database of handwritten digits for a 4-classes problem where the problem was solved significantly faster than the one-versus-three and the one-versus-one real SVM strategies: the computational time taken by the pure complex SVM was almost half that of the real SVMs, but the error rate increased.

### 3. Applications

Now that the main designs of the complex and hypercomplex extensions of SVMs have been reviewed, some important applications of these extensions will be described briefly. Some of the application approaches are limited to using the above section proposals in specific problem solving cases, while others also present contributions (modifications) in the design of the algorithm and/or the kernel that deals with complex and hypercomplex-valued input and/or output data.

#### 3.1. Signal Processing

In [32] the applications of a multi-class SVM and a complex SVM were presented [14] to classify four types of human heartbeat using electrocardiogram signals (ECG) as the input data. The purpose of this application was to aid in the diagnosis of arrhythmia. The complex kernel function presented in [14] was used, and a preprocessing of the ECG beat signal employing a discrete Fourier Transform (DFT) was performed. Accuracies between 86% and 94% were obtained. Additionally, a discussion of the extension of input and output spaces of arbitrary dimensions using Clifford SVM [9] was presented.

The direction of arrival (DOA) estimation problem is a signal processing problem that deals with the obtention of the direction from which a propagating wave arrives at a set of sensors (sensor array). In [33] an approach to solve the DOA problem using a reformulation of one of the most used DOA algorithms, the Minimum Variance Distortionless Response (MVDR) [34] combined with an SVM extension that “can be viewed as a particular case of CSVM [9]” is presented to obtain the SVM-MVDR algorithm. Then, a theoretical relationship (an equivalency) between the sample-based estimation of the output power spectrum of each filter at a certain frequency (SBE) computed by the MVDR and the SBE and estimated by the Multiple Signal Classification (MUSIC) [35,36] is derived by analyzing the high-resolution nonparametric spectrum estimation procedure MVDR. This equivalency is possible when matrices corresponding to the signal space of the MVDR are dropped and the noise eigenvalues are changed by ones. Thus, using the above relationship, an SVM-MUSIC is proposed to solve the DOA estimation problem.

As the performance of the standard MUSIC and MVDR, and the SVM algorithms to solve DOA, are reduced when they deal with coherent signals, the authors proposed the application of another solution for the DOA problem, the spatial smoothing (SS) technique [37] to “provide algorithms with coherent signal detection abilities”. SS divides the sensor array into subarrays, performs individual algorithms in the subarrays, and then calculates the average of the estimated spectra. The experimental results show that the SVM-MUSIC and SVM-MVDR combined with SS obtained the best recognition results compared to their classic versions. The beamforming problem involves the estimation of a signal from a given direction. It is usually defined as the most prominent technique to estimate DOA. In [38], a proposal to solve beamforming by modification of the classic MVDR or Capon beamformer [39,40] is presented, including a regularization term, the so-called  $\epsilon$ -insensitive loss function that penalizes sidelobe levels and allows a certain error in the desired signal array response direction. The result is a convex optimization problem with linear restrictions, which is equal to the SVR problem. Although initially the SVR-MVDR problem was defined in the complex domain, where inputs, outputs, and weights are complex-valued vectors and matrices, this formulation was later transformed from complex to real domains by splitting the real and imaginary parts of the desired beamformer output and, the transformed steering vectors, obtaining a split-definition for the beamformer weights. The proposed method was shown to provide “suitable results” and to outperform Capon and SpheRCB beamformers [40], especially in environments where the signal-to-noise ratio

(SNR) is high, the number of available snapshots is limited, and the calculation of the signal covariance matrix is a difficult task for classic approaches.

Ground penetrating radar (GPR) is a tool that uses radar pulses to image subsurfaces, and it is a very useful sensor in the fields of defense, civil engineering, environment, agriculture, among others. "In civil engineering, the information of the vertical structure of the stratified media can be extracted from radar profiles using echo detection and amplitude estimation. Echo detection provides the time delay estimation (TDE) associated with each interface (which provides important information about the probed media structure), whereas amplitude estimation is used to retrieve the wave speed within each layer [41]." In [41] a combined approach of SVR and forward-backward linear prediction (FBLP) [42] is presented to solving it. The objective function of SVR as formulated in the complex domain using Wirtinger's calculus as it is done in [14]. The experimental results using overlapping and nonoverlapping signals and limited snapshots show that FBLP-SVR outperforms standard FBLP and forward linear prediction (FLP) [43]. Furthermore, FBLP-SVR requires only one snapshot to function. Signals of the geo-positioning system (GPS) and multichannel human heart rate during a very long cycling episode are analyzed in [44] to determine the specific biomedical features, obtain a relation between the heart rate and slopes for downhill and uphill cycling, and the mean heart rate evolution on flat segments. The signals are pre-processed using low-pass finite impulse response (FIR) noise-filtering. In addition, uniform resampling, GPS, and Google maps region information data are merged, and segments of signal patterns are selected to fulfill some criteria (altitude gradient within given limits for uphill, downhill, or flat cycling). Then, the cross-covariance function and the correlation coefficient between data segments are computed to select the data segments that have 95% confidence related to the regression lines. Then a three-feature vector is made for each cycling segment and a two-class classification is performed using NN, complex SVM of Bouboulis et al. [14], and the k-nearest neighbor method, to distinguish between the mean altitude less than 1500 m and greater than 1500 m. The experimental results show that a higher classification accuracy was achieved by NN with sigmoidal transfer functions in the first layer and the probabilistic softmax transfer function in the second layer. The output layer provides a probability for each class based on Bayes' theorem. This is an example of the application of a complex SVM with real-valued data. When the data is not naturally embedded in complex or hypercomplex spaces, the advantages are usually not as great compared to real-valued approaches, such as the NN of this case.

In Table 2, the applications mentioned in this subsection to solve signal processing problems are presented in a compact manner.

**Table 2.** Applications of complex and hypercomplex SVMs for solving signal processing problems.

Ref./Year	Application	SVM Extension Used
[38]/2004	Beamforming problem	Modification of MVDR and a complex-valued SVM.
[33]/2012	DOA estimation problem	CSVM-MVDR [9], SVM-MUSIC, SS-CSVM-MVDR and SS-SVM-MUSIC
[32]/2016	Classification of four types of human heartbeat using ECG signals as the input data	Multi-class complex-valued SVM [14]
[44]/2017	Determination of biomedical features during cycling episodes from GPS and multichannel human heart rate signals	Complex-valued SVM [14]
[41]/2018	Echo detection and amplitude estimation of GPR signals	Combined approach of a complex-valued SVR and FBLP

### 3.2. Pattern Recognition and Classification

In 2016, a survey of advances in image analysis and multi-channel classification for terahertz (THz) pulse imaging and dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) was provided in [45]. The extension of SVM to Clifford SVM [8,9] is analyzed as an image classification method that uses the high-grade elements of the multivector (elements of grade  $> 0$ , i.e., elements that are not the scalar element of the multivector) to describe the multi-channel image, as well as the application of extreme learning machine (ELM).

In [46], the authors show the use of geometrical algebra in the most convenient way to deal with pattern description for higher dimensional spaces. A GA-SVM was applied to classify substances based on their THz spectra, and the results were compared with those obtained using a real-valued SVM. Using the GA framework, it was proved that substances with different thicknesses have coplanar multivectors, which allowed the authors to classify them. The method showed to be dispersion-independent and improved the accuracy and robustness regarding the real-valued SVM.

A modification of the CSVM of [9] involving a multi-output support vector regression (MSVR) model with geometric rotation invariance using GA was proposed in [47]. The main difference between [9] and [47] is that in [9], the primal equation is considered one independent slack variable for each input data multivector and for each multivector component (i.e., one slack variable for each different-grade element of the multivector per each input multivector). This could lead to treating each input sample unfairly, according to the authors of [47], although it is the CSVM key to solving one optimization problem per different-grade element of each input multivector and using the hypercomplex kernel to obtain the direct sum of the linear spaces to obtain multi-classification. Conversely, in the proposal [47], one slack variable is defined for each input multivector data. The experiments show an application of the MSVR to remove clouds in remote-sensing images. A multi-output SVM was used to detect ground objects, thin clouds, and thick clouds. After the detection, the SVM is combined with a support vector value contourlet transform (SVVCT) to achieve multi-scale, multi-direction, and multi-resolution decomposition of remote sensing images. The algorithm presented is effective in removing clouds, but very complex and proportional to the number of decomposed subbands.

Tasks of pattern recognition using iris data and breast cancer data were performed in [48] and respective recognition rates of 96% and 97.8% were obtained by an extension of a CSVM [9]. Object recognition problems using synthetic and real-object data were solved in [9]. Each object was described using a sequence of quaternions, and each quaternion contained the information of 3D key points of the object surface, in such a way that the object shape is entirely represented by a sequence of quaternions. A CSVM with a quaternion-valued Gaussian kernel was used to classify the objects. A six-class of synthetic objects and another six real objects sensed by a stereo camera were recognized, obtaining a minimum of 87.87% in recognition accuracy and a maximum of 96.15% for the synthetic objects in test phase. Meanwhile, 60% as minimum recognition accuracy for real objects and 84% as maximum recognition accuracy were obtained for test phase. In [14] two multiclass classification experiments were conducted using the MNIST database of handwritten digits. In both experiments, the images of each digit were preprocessed using a Fourier transform, and the 100 most significant complex coefficients of each image were used as the input data. In the first experiment, the one-versus-all real-valued SVM was compared with a quaternary complex SVM. The hypercomplex-valued SVM approach outperformed the real one by obtaining an error rate of 3.46% vs. 3.79%. In the second experiment, only the first four digits of the MNIST database were used as it was shown that a complex pair of hyperplanes separate the space of complex numbers into four parts. Thus, the four-class problem is naturally solved in a complex space. A quaternary complex SVM was compared with a one-versus-three real-valued SVM. The error rate for the complex SVM was a little higher (0.866%) than the one obtained with the real-valued SVM (0.721%). "However, the one-versus-three SVM task required about double the time for training, compared to the complexified quaternary SVM." This is

because the real-valued approach solves four optimization problems, while the complex-valued solves two distinct SVM tasks. See Section 2.2.5.

In Table 3, the applications mentioned in this subsection to solve pattern recognition problems are presented in a compact manner.

**Table 3.** Applications of complex and hypercomplex SVMs for solving pattern recognition problems.

Ref./Year	Application	SVM Extension Used
[9]/2010	Object recognition problems using synthetic and real-object data	CSVM of [9]
[48]/2013	Pattern recognition using iris data and breast cancer data	Extension of a CSVM [9]
[47]/2014	Object recognition and remotion of clouds in remote-sensing images	Modification of the CSVM of [9] involving a multi-output support vector regression combined with a support vector value contourlet transform
[14]/2015	Multiclass classification experiments using the MNIST database of handwritten digits.	Complex-valued SVM [14]
[45]/2016	Survey of advances in multi-channel classification for THz pulse imaging and dynamic contrast-enhanced MRI	Clifford SVM [8,9]
[46]/2016	Classification of substances based on their THz spectra	GA-SVM

### 3.3. Time-Series Prediction and Function Approximation

In [23] (Section 2.2.3), using complex-valued SVR called  $\epsilon$ -complex SVR, a function approximation experiment was conducted in which the Laplacian and Gaussian noises were added to a complex function. In the presence of the Laplacian noise, the C-SVR of Bayro et al. [7,8] performed better than the  $\epsilon$ -SVR, the LS-SVR, and the M-SVR. Meanwhile, a function added with the Gaussian noise was better-approximated by the  $\epsilon$ -SVR. In addition, in [23], a chaotic time-series prediction task was solved. With the Lorenz attractor [49] data, the authors derived an eight-step predictor with a  $\epsilon$ -quaternion-valued SVR, and the rotationally invariant loss function defined by their approach allows the production of better models compared to those obtained using the C-SVR.

It is important to note that the solution to the set of equations that models the Lorenz attractor has several applications in physical systems that show chaotic behavior; thus predicting and identifying chaotic systems using complex- and hypercomplex-valued SVRs is a promising application that could lead to obtaining better accuracies than approaches that use real-valued regressors or predictors.

Two experiments in time-series prediction using a recurrent CSVM was presented in [9]. The authors designed a recurrent prediction system that consisted of a long short-term memory (LSTM) [20] neural network connected to a CSVM. The LSTM module solves the problem of identifying temporal dependencies between input and output data; meanwhile, the CSVM provides precision to the prediction system by optimizing the learning of the data mapping. For this, the water levels in the Venice Lagoon during the periods from 1980 to 1989 and 1990 to 1995 were used as the input data. Four-hundred records were used for training and 600 to be predicted by the LSTM-CSVM, the minimum training error was 0.0019. In the second forecasting experiment, the Mackey-Glass time series data was used and the LSTM-CSVM achieved the lower error compared to error of Echo-state networks [50], Evolino [51,52] and LSTM.

In addition, in [9], a case of multiple function approximation was solved using a CSVM. Two different functions involving  $\sin()$  and  $\cos()$  were interpolated at the same time, with a two output CSVM in which 50 points of each function were used to train the CSVM and experiments interpolating 100 and 400 points were performed, obtaining highly accurate values.

In [14], a complex-valued function such as  $\text{sinc}()$  was approximated using a real CSV and a DRC-SVR (see Section 2.2.5). The results obtained shows that the DRC-SVR “fails to capture the complex structure of the function”, and, on the other hand, the CSV estimates the function with a mean MSE that is equivalent to  $-15.75$  dB, vs.  $-10.42$  dB of the DRC-SVR.

In Table 4, the applications mentioned in this subsection for solving time-series prediction and function approximation problems are presented in a compact manner.

**Table 4.** Applications of complex and hypercomplex SVMs for time-series prediction and function approximation problems.

Ref./Year	Application	SVM Extension Used
[23]/2010	Complex function approximation with Laplacian and Gaussian noises and chaotic time-series prediction	$\epsilon$ -SVR of [23] and complex valued SVR of [7,8]
[9]/2010	Time-series prediction using Venice Lagoon and Mackey-Glass datasets	Combined recurrent system of LSTM-CSVM
[9]/2010	Complex function approximation $\text{sinc}()$ and $\text{cosc}()$	CSVM
[14]/2015	Complex function approximation of $\text{sinc}()$	Complex-valued SVM [14]

### 3.4. Frequency Domain Processing, Linear and Nonlinear Filtering

For solving handwritten character recognition using the fast Fourier transform (FFT) coefficients of data from MNIST, fashion MNIST, extended MNIST and Latin OCR benchmark databases a complex-valued neural network with kernel activation functions (KAFs) was presented in [53]. Even though this approach is applied to neural networks instead of SVMs, it is included in this survey to represent one of the proposals to solve one of the main issues of complex or hypercomplex SVMs. It is the design of kernels that process this type of data, and that could be used with all the SVM extensions of the Section 2. Thus, in [53] the ideas of [54–56] are combined to design a widely linear kernel activation function model (WL-KAF). The WL-KAF allows the NN to learn the shape of a non-parametric kernel activation function defined in the complex domain by modeling each activation function “with a small number of complex-valued adaptable parameters, representing the linear coefficients in a kernel-based expansion”, and by using widely linear framework; the model was extended to widely linear kernels with more expressiveness than the classic KAF [54]. This is due to the fact that the classic KAF loses two outputs from the complex kernel by forcing the real and imaginary constraints to be equal, and the cross terms are equal with the scale factor of  $-1$ . The WL-KAF approach applied for solving the frequency domain representation of the handwritten characters achieved better accuracy than the real-valued NN.

Nonlinear channel equalization (NCE) applications using the kernel ridge regression (KRR) with quaternion kernels are shown in [57]. The transmission channel was modeled as a linear filter with a memory-less nonlinearity stage corrupted by noise. Then, a Gaussian quaternion-valued kernel was designed and used in KRR to identify one original message among the noisy measurements. The signals were described as quaternions and, in the comparison analysis performed the quaternion-valued kernel outperformed the real-valued, obtained better accuracy thanks to its ability to capture data geometric and topologic relationships.

In [11,14,23] the linear and nonlinear channel equalization, channel identification and filtering problems were tackled. See Sections 2.2.1, 2.2.2 and 2.2.5.

In [12], the authors consider an application for solving the equalization of a 4-symbol quadrature amplitude modulated (4-QAM) signal over a complex linear communication channel that was added with Gaussian noise, Section 2.2.2.

The frequency estimation problem was solved in [58,59], where a complex-valued support vector autoregressive technique was used. In [59], the autoregressive moving average (ARMA) system

identification and non-parametric spectral analysis were formulated using a framework that allows the authors to identify the Hilbert signal space in the model, define a robust cost function, and minimize a constrained and regularized SVM-like functional with Lagrange multipliers method. Then, a generalized formulation of SVM for dealing with linear signal processing problems was established.

In Table 5, the applications mentioned in this subsection for solving frequency domain processing, and linear and nonlinear filtering problems are presented in a compact manner.

**Table 5.** Applications of complex and hypercomplex SVMs for frequency domain processing, linear and nonlinear filtering problems.

Ref./Year	Application	SVM Extension Used
[59]/2005 and [58]/2011	Frequency estimation problem	Complex-valued support vector autoregressive technique
[11]/2004, [23]/2010 and [14]/2015	Linear and nonlinear channel equalization, channel identification and filtering problems	Complex valued SVMs of [11,14,23]
[12]/2007	Equalization of a 4-symbol quadrature amplitude modulated (4-QAM) signal over a complex linear communication channel that was added with Gaussian noise	Complex valued SVMs
[57]/2014	Nonlinear channel equalization	Kernel ridge regression with quaternion kernels
[53]/2019	Handwritten character recognition using the fast Fourier transform (FFT) coefficients of data from MNIST, fashion MNIST, extended MNIST and Latin OCR benchmark databases	Complex-valued neural network with kernel activation functions

### 3.5. Robotics, Computer Vision

An approach to predict 3D rotation and 3D transformations using computer vision data is presented in [8]. Using a multivector-valued SVM and the GA framework [16,17] the authors solve the problem of rigid motion estimation that is necessary for a robot gripper to move along a 3D curve. With the 3D target position computed from the images of a stereo camera system, the SMVM estimates the motion of the gripper as a sequence of multivectors of the motor algebra  $G_{3,0,1}^+$ . For the problem of 3D pose estimation, a trinocular camera, a triangulation algorithm, and the Hough transform were used to obtain relevant 3D points of an image, from which the 3D lines and their intersection points were computed using a regression to correct the distorted 3D points and to avoid the problem of occluded objects. Once the points, lines, and planes to describe object shapes as multivectors of the motor algebra are represented, they are used to train an SMVM that returns the multivector pose of the described object as two points lying on a 3D line, which were used as reference points on which the robot gripper has to position itself in order to hold the object.

Another proposal for predicting 3D transformations is presented in [60]; they used a technique that approximates the error metric for the blending of unit dual quaternions. The authors used the GA framework to represent the 3D rigid transformation as a unit dual quaternion as in [61]. Then, with observations that are pairs of joint angles and rigid transformations of the fingertip relative to its root, regressors were constructed to predict the position and orientation of a fingertip with two rotational joints without any knowledge of its kinematic model. The authors also proved the effectiveness of their method in predicting the elastic deformation markers positions on a balloon as it is squashed, constructing the regressors of 3D rigid transformations of the marker and using those regressors as the 3D shape of the balloon. The regressors used were DQ regression, SVR, and ridge regression.

In [57], KRR was used to simulate 3D inertial body sensor data. This is usually used in robotics to teach humanoid robots motion and movement. The paper explores “the existence and uniqueness conditions of quaternion reproducing kernel Hilbert spaces (QRKHS)” to aid in the designing of the

quaternion-valued kernels that are needed in hypercomplex SVMs and SVRs. Examples of quaternion valued-kernels were designed: the quaternion cubic kernel and the quaternion-valued Gaussian kernel. The first kernel mentioned was used in KRR to perform a one-step-ahead prediction of the limb trajectory in Tai Chi sequences, and a comparison of the results between the KRRs was performed using real, vector, and quaternion cubic kernels. The quaternion-valued kernels were shown to obtain lower mean squared errors than their real-valued counterparts. The authors argue that it was done because the hypercomplex-valued kernels “capture the inherent data relationships in a more accurate and physically-meaningful way, while being robust to overfitting”.

The real- and quaternion-valued Gaussian kernels were used in KRR to perform channel equalization. See Section 3.4.

Even when an SVR was not used in the above two applications, quaternion-valued kernels were designed and proved in real applications and their advantages over real-valued kernels were demonstrated. The advantages of using quaternion-valued kernels could be extended when SVR is used instead of KRR.

Another technique that can be used to teach robots human movement and motion is the one presented in [13] and revised in Section 2.2.4. Here, an octonionic SVR was applied to the biomechanical study of human locomotion, called the gait analysis. The objective for the regressor was to estimate correct measurements of key events of the endpoint foot trajectory during human treadmill walking. The task was achieved with high accuracy.

All the above proposals to predict 3D transformations could be used in robotics, computer vision and computer graphics to model motion and 3D shapes, simulation of the interaction between robots and their environment, predictive control, and structure from motion, among others. The robot path-planning problem was addressed in [9]. This problem was solved using an LSTM-CSVM recurrent system and training the LSTM module with reinforcement learning (RL). A labyrinth environment was modeled as an occupation grid. Each observation of the navigator robot was described using a quaternion, where each element of the quaternion indicates if one robot sensor reads 0, and if the next cell of the grid was free or 1 if it was occupied by an obstacle. A cross-neighborhood, as regards the actual position of the robot on the grid, was used; thus, four inputs of one quaternion were used to describe an observation (forward, backward, left, and right), and two positions of a second quaternion were used for determining the 2D coordinates of the position of the robot on the plane. The four outputs were packed in a quaternion, and they represented the action that the robot has to perform in the actual state to get the maze output (go forward, backward, left, or right). In order for the path-planning system to be used, firstly, this had to be trained using some mazes, and then proven with never-before-seen mazes. Therefore, once the path is computed, the robot can follow it in real time. Here, the hypercomplex inputs and outputs were used to get a MIMO SVM that reads two quaternions and returns one. See Section 2.1.

In Table 6, the applications mentioned in this subsection for solving robotics and computer vision problems are presented in a compact manner.

**Table 6.** Applications of complex and hypercomplex SVMs for robotics and computer vision problems.

Ref./Year	Application	SVM Extension Used
[8]/2001	Prediction of 3D rotation and 3D transformations using computer vision data	Multivector-valued SVM and the GA framework
[9]/2010	Robot path-planning	Combined recurrent system of LSTM-CSVM
[13]/2012	Study of human locomotion (the gait analysis)	Complex valued SVR
[57]/2014	Simulation of 3D inertial body sensor data	KRR using quaternion-valued kernels
[60]/2017	Prediction of 3D transformations	DQ regression, SVR and ridge regression to approximate the error metric for the blending of unit dual quaternions

#### 4. Conclusions and Prospective Work

In this survey, several proposals were reviewed to extend the Support Vector Machine algorithm to deal with complex and hypercomplex-valued inputs, outputs, and kernels. The first attempts to achieve this consisted of splitting the complex and hypercomplex numbers into real and complex (hypercomplex) parts, and then solving the optimization problems independently in the real domain. Then, it was proven that this approach loses the benefits of the complex (hypercomplex) spaces, such as sparsity and geometric and topologic features that the vectors (multivectors) have when are embedded in their spaces. All the proposals that exploit these features by solving the SVM optimization problem in its complex (hypercomplex) space were proven to achieve higher accuracies than those that solve the problem using splitting procedures.

The most complete and general extensions of SVMs, that deal with complex/hypercomplex-valued data are those that, besides considering input and output data that is complex/hypercomplex valued, define the optimization problem (primal and dual problems) and the kernel using complex and hypercomplex algebras and calculus frameworks, such as Clifford and GA, widely linear analysis and Wirtinger's calculus; although it was also proved [14] that the design of a complex/hypercomplex-valued kernel that maps the input data onto a complex/hypercomplex RKHS is the key to obtaining the benefits of solving the problem in these spaces, such as the preservation of the geometric and topological special features of these spaces.

Therefore, for dealing with data that is naturally embedded in complex (hypercomplex) spaces the approaches that were presented in the Section 2 of this paper are the best choices to obtain higher accuracies and lower run time in problems of classification and regression.

Another important conclusion is about the compatibility of the complex numbers and their frameworks with the wave nature due to the waves have intensity and phase, so two numbers are needed to describe it.

In the applications section, it is shown that the extensions of SVM that were designed using complex and hypercomplex mathematical tools can solve efficiently and effectively real-world problems that involve the processing and classification, interpolation, or regression of complex signals, frequency domain data, ECG signals, MRI data, chaotic time series, complex/hypercomplex forecasting, approximation of complex/multiple functions, complex beamforming and DOA, GPR and GPS data, linear and nonlinear complex filtering, pattern recognition, robotics, and computer vision, among others.

Regarding the prospective of complex/hypercomplex extensions and applications of SVMs, one promising development is wave informatics because, as it was said before, the natural compatibility between complex numbers and waves. The applications are important and numerous as they can be seen in Section 3.

Another area of opportunity for researching is the processing, classification, regression, prediction, identification, and simulation of chaotic systems. All the deterministic aperiodic system that is sensitive to the initial conditions is called a chaotic system. Their study is very important as some chaotic systems examples are the weather, some electrocardiograms, and encephalograms time series, the stock market time series, and fractals, among others. These systems are also naturally modeled using complex-valued functions, and therefore, the extensions of complex-valued SVMs are the best-suited to process them.

As was mentioned, perhaps the most general extension of complex/hypercomplex valued SVMs is the one presented in [9]. Here, the use of GA framework allowed the authors to define an SVM problem that processes complex/hypercomplex-valued inputs, outputs, and kernels. Nevertheless, the design of complex/hypercomplex kernels remains a main open issue to explore the extensions of SVM to higher-dimensional hypercomplex spaces (more than quaternions, dual quaternions, or octonions). Thus, exploring the definition of high-dimensional Clifford and/or GA to design hypercomplex kernels, and to exploit the benefits of the sparsity of these vector spaces and their power of geometric expressiveness is a promising research area. Even in hypercomplex high-dimensional Clifford and/or



GA, the design of kernels may not be necessary because the sparsity could make it possible to obtain linear functions to correctly separate the input data, reducing the complexity of the SVM algorithm in both training and testing stages.

This exploration of higher-dimensional hypercomplex GA could benefit the application of hypercomplex SVMs to simulate another main issue of chaos theory: the geometry of fractals and even the identification, modeling and control of chaotic dynamic systems. The geometry of fractals is defined as the geometry of nature, chaos, and order. The geometric shape of fractals describes sinuous curves, spirals, and filaments that twist about themselves, giving elaborate figures whose details are lost in infinity. Therefore, using very higher-dimensional GA, fractals could be described as nonlinear combinations of the highly complex geometric entities that can be defined in that higher-dimension GA as multivectors, and a hypercomplex-valued SVM could approximate those multivectors. Now, the problem of exploring these GA seems to be a computationally expensive and complex task, but the emergence of another new area of research could make it possible: quantum computing.

Quantum computing stands as another prospective area of research and application of complex and hypercomplex-valued SVM and as an opportunity to develop any type of complex/hypercomplex-valued machine learning techniques [62–67]. Quantum computing is a new computational paradigm that uses the quantum theory to develop computer technology. In this computing technology, the quantum system qubit is used instead of the classic computational paradigm’s bit. Quantum systems are those that exhibit both particle and wave-like behavior. Therefore, modeling this dual behavior using complex numbers would seem to be natural, wherein the real part could be used to represent particle behavior and the imaginary part to describe wave phase that gives rise to the interference patterns. Quantum computing has made it possible to solve some problems that cannot be solved using classic computing, such as the factoring of integers and discrete logarithm computation. In this paper, it has been shown that when data of a problem fits well with the wave nature of complex (hypercomplex) numbers, it is better to use an algorithm that works in the vector space in which the fitted data is defined in order to take advantage of their geometric relationships and distributions. Hence, the reviewed extensions of SVMs are suited to deal with the basic unit of quantum information, the qubit.

In Figure 1 several prospective and promising works to continue developing theory and applications of complex and hypercomplex SVMs are illustrated.

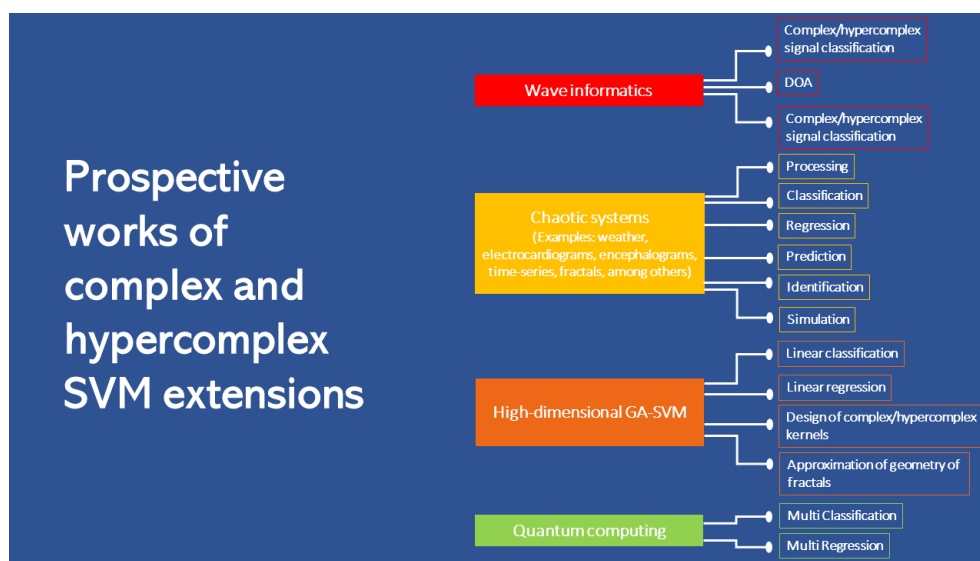


Figure 1. Prospective works of complex and hypercomplex SVM extensions.

**Funding:** This research was funded by CONACYT México grants numbers CB256769, CB258068 and PN-4107.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
2. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: Berlin, Germany, 1995.
3. Bartlett, P.L. The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **1998**, *44*, 525–536. [[CrossRef](#)]
4. Shawe-Taylor, J.; Bartlett, P.L.; Williamson, R.C.; Anthony, M. Structural Risk Minimization over Data-dependent Hierarchies. *IEEE Trans. Inf. Theor.* **1998**, *44*, 1926–1940, doi:10.1109/18.705570. [[CrossRef](#)]
5. Shawe-Taylor, J.; Cristianini, N. Margin distribution and soft margin. In *Advances in Large Margin Classifiers*; Smola, A.J., Ed.; The MIT Press: Cambridge, MA, USA, 2000; pp. 349–358.
6. Top 100 SVM Publications. Available online: <http://www.svms.org/top100.html> (accessed on 30 March 2019).
7. BayroCorrochano, E.; Vallejo, R. Geometric neural networks and support multivector machines. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IEEE IJCNN 2000), Como, Italy, 27 July 2000; pp. 389–394.
8. Bayro-Corrochano, E.; Vallejo, R. SVMs using geometric algebra for 3D computer vision. In Proceedings of the International Joint Conference on Neural Networks (IJCNN'01), Washington, DC, USA, 15–19 July 2001; Proceedings (Cat. No.01CH37222); Volume 2, pp. 872–877, doi:10.1109/IJCNN.2001.939474. [[CrossRef](#)]
9. BayroCorrochano, E.J.; AranaDaniel, N. Clifford Support Vector Machines for Classification, Regression, and Recurrence. *IEEE Trans. Neural Netw.* **2010**, *21*, 1731–1746, doi:10.1109/TNN.2010.2060352. [[CrossRef](#)] [[PubMed](#)]
10. PérezCruz, F.; Camps, G.; Soria, E.; Pérez, J.; FigueirasVidal, A.R.; ArtésRodríguez, A. Multidimensional function approximation and regression estimation. In *International Conference on Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2002.
11. SanchezFernandez, M.; de PradoCumplido, M.; ArenasGarcia, J.; PerezCruz, F. SVM multiregression for nonlinear channel estimation in multipleinput multipleoutput systems. *IEEE Trans. Signal Process.* **2004**, *52*, 2298–2307. [[CrossRef](#)]
12. Shilton, A.; Lai, D. Quaternionic and complexvalued support vector regression for equalization and function approximation. In Proceedings of the International Joint Conference on Neural Networks IJCNN 2007, Orlando, FL, USA, 12–17 August 2007; pp. 920–925.
13. Shilton, A.; Lai, D.; Santhiranayagam, B.; Palaniswami, M. A note on octonionic support vector regression. *IEEE Trans. Syst. Man Cybern.* **2012**, *42*, 950–955. [[CrossRef](#)] [[PubMed](#)]
14. Bouboulis, P.; Theodoridis, S.; Mavroforakis, C.; Evaggelatos-Dalla, L. Complex Support Vector Machines for Regression and Quaternary Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 1260–1274, doi:10.1109/TNNLS.2014.2336679. [[CrossRef](#)]
15. López-González, G.; Arana-Daniel, N.; Bayro-Corrochano, E. Quaternion support vector classifier. *Intell. Data Anal.* **2016**, *20*, doi:10.3233/IDA-160849. [[CrossRef](#)]
16. Hestenes, D.; Sobczyk, G. *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics*; Springer Science and Business Media: Berlin, Germany, 2012; Volume 5.
17. Perwass, C. *Geometric Algebra with Applications in Engineering*; Springer: Berlin/Heidelberg, Germany, 2009.
18. Goto, E. *The Parametron—A New Circuit Element which Uses Non-Linear Reactors*; Paper of Technical Group of Electronic Computers and Nonlinear Theory; IECE, 1954. (In Japanese)
19. Goto, E. On the application of parametrically excited non-linear resonators. *J. Inst. Electr. Commun. Eng. Jpn. IECE* **1955**, *38*, 2761. (In Japanese)
20. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780, doi:10.1162/neco.1997.9.8.1735. [[CrossRef](#)]
21. López-González, G.; Arana-Daniel, N.; Bayro-Corrochano, E. Parallel Clifford Support Vector Machines Using the Gaussian Kernel. *Adv. Appl. Clifford Algebras* **2016**, doi:10.1007/s00006-016-0726-2. [[CrossRef](#)]
22. Dempe, S. *Foundations of Bilevel Programming*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2002.
23. Shilton, A.; Lai, D.; Palaniswami, M. A division algebraic framework for multidimensional support vector regression. *IEEE Trans. Syst. Man Cybern.* **2010**, *40*, 517–528. [[CrossRef](#)] [[PubMed](#)]
24. Suykens, J.A.K.; Gestel, T.V.; Brabanter, J.D.; Moor, B.D.; Vandewalle, J. *Least Squares Support Vector Machines*; World Scientific: Singapore, 2002.

25. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
26. Huang, G.B.; Chen, L.; Siew, C.K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Netw.* **2006**, *17*, 879–892. [[CrossRef](#)] [[PubMed](#)]
27. Weston, J.; Chapelle, O.; Elisseeff, A.; Vapnik, V. *Kernel Dependency Estimation*; Technical Report 098; Max Planck Institute for Biological Cybernetics: Tübingen, Germany, 2002.
28. Santhiranaayagam, B.K.; Lai, D.T.H.; Begg, R.K.; Palaniswami, M. Correlations between end point foot trajectories and inertial sensor data. In Proceedings of the 2010 Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Brisbane, Australia, 7–10 December 2010; pp. 315–320.
29. Picinbono, B.; Chevalier, P. Widely linear estimation with complex data. *IEEE Trans. Signal Process.* **1995**, *43*, 2030–2033, doi:10.1109/78.403373. [[CrossRef](#)]
30. Kracht, M.; Kreyszig, E. *Methods of Complex Analysis in Partial Differential Equations and Applications*; John Wiley and Sons: New York, NY, USA, 1988.
31. Bouboulis, P.; Theodoridis, S. Extension of Wirtinger’s calculus to reproducing kernel Hilbert spaces and the complex kernel LMS. *IEEE Trans. Signal Process.* **2011**, *59*, 964–978. [[CrossRef](#)]
32. Jannah, N.; Hadjiloucas, S. Detection of ECG arrhythmia conditions using CSVM and MSVM classifiers. In Proceedings of the 2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 12 December 2015; pp. 1–2, doi:10.1109/SPMB.2015.7405453. [[CrossRef](#)]
33. El Gonnouni, A.; Martinez-Ramon, M.; Rojo-Alvarez, J.L.; Camps-Valls, G.; Figueiras-Vidal, A.R.; Christodoulou, C.G. A Support Vector Machine MUSIC Algorithm. *IEEE Trans. Antennas Propag.* **2012**, *60*, 4901–4910, doi:10.1109/TAP.2012.2209195. [[CrossRef](#)]
34. Bienvenu, G.; Kopp, L. Optimality of high resolution array processing using the eigensystem approach. *IEEE Trans. Acoust. Speech Signal Process.* **1983**, *31*, 1235–1248, doi:10.1109/TASSP.1983.1164185. [[CrossRef](#)]
35. Schmidt, R.O. Multiple emitter location and signal parameter estimation. In Proceedings of the RADC Spectral Estimation Workshop, Griffiss AFB, NY, USA, 3–5 October 1979; pp. 243–258.
36. Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280, doi:10.1109/TAP.1986.1143830. [[CrossRef](#)]
37. Shan, T.-J.; Wax, M.; Kailath, T. On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Trans. Acoust. Speech Signal Process.* **1985**, *33*, 806–811, doi:10.1109/TASSP.1985.1164649. [[CrossRef](#)]
38. Gaudes, C.C.; Via, J.; Santamaria, I. Robust array beamforming with sidelobe control using support vector machines. In Proceedings of the IEEE 5th Workshop on Signal Processing Advances in Wireless Communications, Lisboa, Portugal, 11–14 July 2004; pp. 258–262, doi:10.1109/SPAWC.2004.1439244. [[CrossRef](#)]
39. Trees, H.L.V. *Detection Estimation and Modulation Theory, Part, Part IV, Optimum Array Processing*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2001.
40. Li, J.; Stoica, P.; Wang, Z. On robust Capon beamforming and diagonal loading. *IEEE Trans. Signal Process.* **2003**, *51*, 1702–1715, doi:10.1109/TSP.2003.812831. [[CrossRef](#)]
41. Pan, J.; Le Bastard, C.; Wang, Y.; Sun, M. Time-Delay Estimation Using Ground-Penetrating Radar With a Support Vector Regression-Based Linear Prediction Method. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2833–2840, doi:10.1109/TGRS.2017.2784567. [[CrossRef](#)]
42. Chen, Y.H.; Chiang, C.-T. Kalman-based spatial domain forward-backward linear predictor for DOA estimation. *IEEE Trans. Aerosp. Electron. Syst.* **1995**, *31*, 474–479, doi:10.1109/7.366330. [[CrossRef](#)]
43. Haykin, S. *Adaptive Filter Theory*, 4th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2002.
44. Procházka, A.; Vaseghi, S.; Charvátová, H.; Ťupa, O.; Vyšata, O. Cycling Segments Multimodal Analysis and Classification Using Neural Networks. *Appl. Sci.* **2017**, *7*, 581, doi:10.3390/app7060581. [[CrossRef](#)]
45. Yin, X.X.; Zhang, Y.; Cao, J.; Wu, J.L.; Hadjiloucas, S. Exploring the complementarity of THz pulse imaging and DCE-MRIs: Toward a unified multi-channel classification and a deep learning framework. *Comput. Methods Programs Biomed.* **2016**, *137*, 87–114, doi:10.1016/j.cmpb.2016.08.026. [[CrossRef](#)] [[PubMed](#)]
46. Zhou, S.; Valchev, D.G.; Dinovitser, A.; Chappell, J.M.; Iqbal, A.; Ng, B.W.; Kee, T.W.; Abbott, D. Terahertz Signal Classification Based on Geometric Algebra. *IEEE Trans. Terahertz Sci. Technol.* **2016**, *6*, 793–802, doi:10.1109/TTHZ.2016.2610759. [[CrossRef](#)]
47. Hu, G.; Sun, X.; Liang, D.; Sun, Y. Cloud removal of remote sensing image based on multi-output support vector regression. *J. Syst. Eng. Electron.* **2014**, *25*, 1082–1088, doi:10.1109/JSEE.2014.00124. [[CrossRef](#)]

48. Wang, J.; Zhang, C.; Lu, H.; Hong, W. Geometric algebra support vector machine for pattern recognition. *ICIC Express Lett.* **2013**, *7*, 1031–1035.
49. Lorenz, E. Deterministic nonperiodic flow. *J. Atmos. Sci.* **1963**, *20*, 130–141. [[CrossRef](#)]
50. Jaeger, H.; Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **2004**, *304*, 78–80. [[CrossRef](#)]
51. Schmidhuber, J.; Wierstra, D.; Gomez, F.J. Evolino: Hybrid neuroevolution optimal linear search for sequence prediction. In Proceedings of the 19th International Joint Conference on Artificial Intelligence IJCAI, Edinburgh, UK, 30 July–5 August 2005.
52. Schmidhuber, J.; Gagliolo, M.; Wierstra, D.; Gomez, F. Evolino for Recurrent Support Vector Machines. *arXiv* **2005**, arXiv:cs/0512062.
53. Scardapane, S.; Vaerenbergh, S.V.; Comminiello, D.; Uncini, A. Widely Linear Kernels for Complex-Valued Kernel Activation Functions. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019), Brighton, UK, 12–17 May 2019.
54. Scardapane, S.; Vaerenbergh, S.V.; Hussain, A.; Uncini, A. Complex-valued neural networks with nonparametric activation functions. *IEEE Trans. Emerg. Top. Comput. Intell.* **2018**. [[CrossRef](#)]
55. Scardapane, S.; Vaerenbergh, S.V.; Totaro, S.; Uncini, A. Kafnets: Kernel-based non-parametric activation functions for neural networks. *Neural Netw.* **2019**, *110*, 19–32, doi:10.1016/j.neunet.2018.11.002. [[CrossRef](#)] [[PubMed](#)]
56. Boloix-Tortosa, R.; Murillo-Fuentes, J.J.; Santos, I.; Pérez-Cruz, F. Widely Linear Complex-Valued Kernel Methods for Regression. *IEEE Trans. Signal Process.* **2017**, *65*, 5240–5248, doi:10.1109/TSP.2017.2726991. [[CrossRef](#)]
57. Tobar, F.A.; Mandic, D.P. Quaternion Reproducing Kernel Hilbert Spaces: Existence and Uniqueness Conditions. *IEEE Trans. Inf. Theory* **2014**, *60*, 5736–5749, doi:10.1109/TIT.2014.2333734. [[CrossRef](#)]
58. Xu, N. Applications of Support Vector Machines in Electromagnetic Problems. Ph.D. Thesis, The University of New Mexico, Albuquerque, NM, USA, 2011.
59. Rojo-Álvarez, J.; Camps-Valls, G.; Martínez-Ramón, M.; Soria-Olivas, E.; Navia-Vázquez, A.; Figueiras-Vidal, A. Support vector machines framework for linear signal processing. *Signal Process.* **2005**, *85*, 2316–2326, doi:10.1016/j.sigpro.2004.12.015. [[CrossRef](#)]
60. Funatomi, T.; Iiyama, M.; Kakusho, K.; Minoh, M. Regression of 3D rigid transformations on real-valued vectors in closed form. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 6412–6419, doi:10.1109/ICRA.2017.7989757. [[CrossRef](#)]
61. Kavan, L.; Collins, S.; Žára, J.; O’Sullivan, C. Geometric Skinning with Approximate Dual Quaternion Blending. *ACM Trans. Graph.* **2008**, *27*, 105:1–105:23, doi:10.1145/1409625.1409627. [[CrossRef](#)]
62. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed.; Cambridge University Press: New York, NY, USA, 2011.
63. Rieffel, E.; Polak, W. *Quantum Computing: A Gentle Introduction*, 1st ed.; The MIT Press: Cambridge, MA, USA, 2011.
64. Yanofsky, N.S. An Introduction to Quantum Computing. *arXiv* **2007**, arXiv:0708.0261.
65. Gupta, S.; Mohanta, S.; Chakraborty, M.; Ghosh, S. Quantum machine learning-using quantum computation in artificial intelligence and deep neural networks: Quantum computation and machine learning in artificial intelligence. In Proceedings of the 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), Bangkok, Thailand, 16–18 August 2017; pp. 268–274, doi:10.1109/IEMECON.2017.8079602. [[CrossRef](#)]
66. Wittek, P. *Quantum Machine Learning: What Quantum Computing Means to Data Mining*; Academic Press: Cambridge, MA, USA, 2014.
67. Schuld, M.; Sinayskiy, I.; Petruccione, F. An introduction to quantum machine learning. *Contemp. Phys.* **2014**, *56*, doi:10.1080/00107514.2014.964942. [[CrossRef](#)]

