

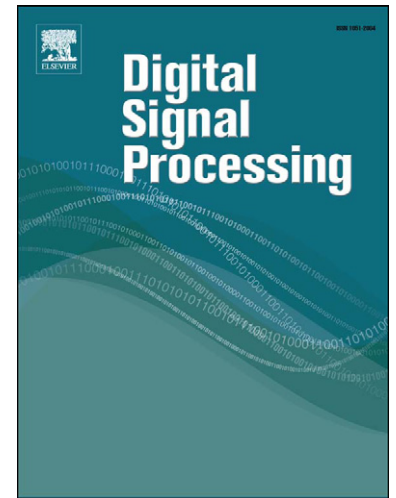
Accepted Manuscript

Application of Complex Extreme Learning Machine to Multiclass Classification problems with High Dimensionality: a THz spectra classification problem

X.-X. Yin, S. Hadjiloucas, J. He, Y. Zhang, Y. Wang, D. Zhang

PII: S1051-2004(15)00042-1
DOI: <http://dx.doi.org/10.1016/j.dsp.2015.01.007>
Reference: YDSPR 1718

To appear in: *Digital Signal Processing*



Please cite this article in press as: X.-X. Yin et al., Application of Complex Extreme Learning Machine to Multiclass Classification problems with High Dimensionality: a THz spectra classification problem, *Digital Signal Process.* (2015), <http://dx.doi.org/10.1016/j.dsp.2015.01.007>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Application of Complex Extreme Learning Machine to Multiclass Classification problems with High Dimensionality: a THz spectra classification problem

X.-X. Yin ^a, S. Hadjiloucas ^b, J. He ^a, Y. Zhang ^a, Y. Wang ^c,
and D. Zhang ^d

^aCenter for Applied Informatics and College of Engineering and Science, Victoria University, VIC 8001, Australia.

^bSchool of Systems Engineering, University of Reading, Reading RG6 6AY, U.K.

^cDepartment of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA

^dDepartment of Computing, the Hong Kong Polytechnic University, Kowloon, Hong Kong, China

Abstract

We extend extreme learning machine (ELM) classifiers to complex Reproducing Kernel Hilbert Spaces (RKHS) where the input/output variables as well as the optimisation variables are complex-valued. A new family of classifiers, called complex-valued ELM (CELM) suitable for complex-valued multiple- input-multiple-output processing is introduced. In the proposed method, the associated Lagrangian is computed using induced RKHS kernels, adopting a Wirtinger calculus approach formulated as a constrained optimization problem similarly to the conventional ELM classifier formulation. When training the CELM, the Karush-Khun-Tucker (KKT) theorem is used to solve the dual optimization problem that consists of satisfying simultaneously smallest training error as well as smallest norm of output weights criteria. The proposed formulation also addresses aspects of quaternary classification within a Clifford algebra context. For 2D complex-valued inputs, user-defined complex-coupled hyper-planes divide the classifier input space into four partitions. For 3D complex-valued inputs, the formulation generates three pairs of complex-coupled hyper-planes through orthogonal projections. The six hyper-planes then divide the 3D space into eight partitions. It is shown that the CELM problem formulation is equivalent to solving six real-valued ELM tasks, which are induced by projecting the chosen complex kernel across the different user-defined coordinate planes. A classification example of powdered samples on the basis of their terahertz spectral signatures is used to demonstrate the advantages of the CELM classifiers compared to their SVM counterparts. The proposed classifiers retain the advantages of their ELM counterparts, in that they can perform multiclass classification with lower computational complexity than SVM classifiers. Furthermore, because of their ability to perform classification tasks fast, the proposed formulations are of interest to real-time applications.

Key words: complex extreme learning machine, Reproducing Kernel Hilbert Space, quaternary classification, Lagrangian, multiclass classification

Introduction

Feature extraction using kernel based learning [1] and support vector machine (SVM) methodologies reside at the core of a range of interdisciplinary challenges. Their formulation shares concepts from different disciplines such as: linear algebra, mathematics, statistics, signal processing, systems and control theory, optimization, machine learning, pattern recognition, data mining and neural networks. This paper extends the formulation of a very important class of

recently developed classifiers called Extreme Learning Machines (ELMs) to complex-valued problems [2, 3]. The motivation for the proposed extension stems from the fact that the real-valued ELM has shown some of the lowest training errors among machine learning algorithms and in particular support vector machines classifiers (SVMs) [4–7]. ELM classifiers have important applications in information science, image and vision, pattern recognition, robotics, cellular automata, as multi-state inference engines in Boolean networks for bioinformatics, quantum computing, cryptography, communications, as well as particle physics modeling (enabling fermionic or bosonic particle classification [8] on

* X.-X. Yin, Email: xiaoxia.yin@vu.edu.au

the basis of both energy and momentum). Since many physical problems described by wave theory are complex-valued (with their parameters joined by the well-known Kramers-Kronig relationships [9] which relate amplitude and phase), by extending ELMs to complex inputs, their applications domain can dramatically increase, encompassing all types of research associated to the study of the interaction of matter with waves, and in particular spectroscopy (acoustic, dielectric, optical, terahertz, infrared, electron-spin resonance, nuclear magnetic or paramagnetic resonance, etc.) as well as imaging and tomography modalities encountered across the Physical, Chemical and Biomedical disciplines. As a consequence, the proposed extension is fundamental both from a Machine Learning as well as from a Chemometrics perspective [10]. Because the above relations are also analogous to the blurring function (relating amplitude and phase) developed by Bode to describe the dynamics of physical systems, such extension has a wide range of applications across all physical sciences.

ELM classifiers benefit from milder optimization constraints compared to least-squares support vector machines (LS-SVM) and proximal support vector machines (PSVM), lower computational complexity and ability to approximate any target continuous function and classify any disjoint regions with much better (for multi-class cases) generalization performance at much faster learning rates (with potentially thousand-fold improvement) [2, 3]. This remarkable improvement in learning rate makes them particularly useful in real-time applications. Other important features that they have are a) their adoption of a generalized single-hidden-layer feed-forward network (SLFN) whose hidden layer need not be tuned and b) the fact that the optimization aims to achieve not only the smallest training error but also the smallest norm of output weights. These observations are discussed to more detail in the seminal work reported in [11], where it was first established that the SLNF is not necessary neuron-like.

In the current contribution, the classification of higher dimensional input data in complex reproducing kernel Hilbert spaces using ELM algorithms can be viewed as a linear combination of several rational multi-class classification problems with 2D input data. Such approach is further supported by the work by Bouboulis et. al. [12], where within a SVM context, they showed that a derived 2D complex kernel is equivalent to an induced real kernel, formed as a linear combination of two identical 2D real value kernels. A widely linear estimation processing approach is adopted and the sum of real and imaginary parts is employed to model the output weights that connect the hidden layer with the feature mapping of the input into the hidden-layer feature space. For 2D inputs, the coupled complex hyper-planes are used to divide the input space into four partitions. This approach, for example is most appropriate when one wishes to combine spatial and temporal information using the same classifier rules. Because inter-relations of the data at the input space are retained, the proposed approach is expected to lead to improved classifier performance compared with

its real-valued ELM counterpart. To further clarify this concept, a linear estimation processing example is considered for illustrations purposes. Using a classification example of terahertz (THz) powdered samples of various composition, a goal is set to separate them on the basis of their spectral signatures. Inputs are then considered as a sum of two parts (real and imaginary) and the output weights connecting the hidden layer and the feature mapping of the input into the hidden-layer feature space will be obtained with the new classifier. It is our expectation that compared to a real-valued ELM classification scheme, CELM will show increased classification accuracy due to the higher dimensionality of its input space. The proposed scheme should also avoid over-fitting problems. The “over-fitting problem” in this context means that the features from both real part and imaginary part are not fully independent. An advantage of using THz transient datasets to showcase the new algorithm is that we can record the complex insertion loss of the material, which has a real part associated to impedance mismatch of the EM propagating wave at the air-sample interfaces and an absorption part which is associated with the complex component of the wave. By presenting these signatures to the classifier separately we maintain a capability to differentiate between samples according to the origin of the observed difference so there is no loss of information as would be the case of presenting them directly as mixed entries. **This becomes important for example, when the sample is fibrous or striated, since adjacent pixels in one direction will show a high degree of similarity whereas in the other direction a lower degree of similarity. In classifying proteins, and biomedical tissue (muscle, nerve), this can be rather important so we would need to input to the classifier x, y direction pixels separately, each one containing both amplitude and phase information encoding.** Furthermore, since there are two classifiers with their respective input and output spaces, each one can be tuned separately, without compromising its fidelity as would be the case if complex entries were used, where it would be necessary to create a joint input-output map.

The proposed CELM formulation may be seen as a “generalized” single-hidden-layer feed-forward network (SLNF). When viewing CELM as an extension of ELM, the dimension of complex coupled output weights that link the hidden layer nodes is given before the learning processing of the complex inputs, therefore the computational costs are low. As a result, the CELM should have a fast convergence to the desired solution. Furthermore, by deriving the structure of a rational multiclass classifier of high dimensional input data as a linear combination of complex RKHS one can consider to classify complex-valued input datasets with high dimensionality.

On the basis of the above extension to complex kernels, it is also possible to consider an even more feature rich input space assuming multiple projections of the data, by using geometric algebras, such as Clifford algebras in the problem formulation [13, 14]. Geometric algebras (GAs) belong

to the group of associative algebras, and have been an active research area in machine learning and pattern identification since the 1990s, especially within the context of spatio-temporal data processing [15, 16]. They have already led to important discoveries in areas as diverse as biological sciences [17], biomedicine [18], and communications [19]. There are already several examples of such implementations e.g., Clifford neural networks (NN) [20], Clifford support vector machines (SVMs) [21, 22]. All these algorithms show excellent performance compared to their real-valued counterparts. A review of Clifford machine learning applications can be found in [16]. The advantage of adopting a GA framework is that one can solve mathematical problems in a vector space, where the same general algebraic framework may be adopted throughout the calculations by simply extending the computation methods chosen for the pattern identification algorithms to their equivalent vector space. This way, one can retain some information within the structure in the data at all times, without having to modify the algorithm each time the input feature space changes. In the extension of the formulation to 3D inputs for example, as would be the case if amplitude, phase and polarization information were to be presented simultaneously at the input layer of the classifier, one obtains three pairs of coupled complex hyper-planes through orthogonal projections. The six hyper-planes divide the 3D space into eight portions. With the proposed extension, nonlinear estimation may be performed by mapping the input space to an infinite dimensional feature space, where the output is a linear combination of the feature maps.

A major motivation for using two decoupled real ELMs of dimension \mathfrak{D} ($\mathfrak{D} \in \mathbf{R}$) to represent complex-valued ELMs is that they capture the “interrelation” among features in different mode of the matrix format, which are normally hard to capture using a real-valued ELM of dimension $2\mathfrak{D}$. When making the conversion using two decoupled real ELMs of dimension \mathfrak{D} , we not only search the difference between features of \mathfrak{D} dimension, but also search to find the relations among features. Such relations can often exist between features and should not be masked. However, if we simply use an ELM of dimension $2\mathfrak{D}$, we can only find the difference, as it is to assume that the feature of dimension $2\mathfrak{D}$ is fully independent. It is important for a classifier that not only has the ability to classify the difference in features, but also finds the inter-relations among features.

The advantage of the proposed approach is that it is compatible with Wirtinger’s calculus and potentially Clifford algebra classifiers thus enabling the fusion of datasets acquired through different measurement modalities. Although this aspect is not investigated explicitly in the current work, it is of relevance to the Machine Learning community. In principle, a complex kernel could also be used, this was not adopted because of the above mentioned limitation.

There is also a physically intuitive interpretation to the proposed approach. For THz transients, we can record the complex insertion loss of the material, which has a real

part associated to impedance mismatch of the EM propagating wave at the air-sample interfaces and an absorption part which is associated with the complex component of the wave. By presenting these signatures to the classifier separately we maintain a capability to differentiate between samples according to the origin of the observed difference so there is no loss of information as would be the case of presenting them directly as mixed entries. Furthermore, since there are two classifiers with their respective input and output spaces, each one can be tuned separately, without compromising its fidelity as would be the case if complex entries were used, it would be necessary to create a joint input-output map.

It is worth noting that the proposed CELM is different from other works in the literature discussing fully Complex-valued ELM Classifiers [23, 24]. In the case of the first paper mentioned, the authors proposed a hyperbolic secant function to be employed as a fully complex-valued activation function (a kernel function), this differs from the current approach. Furthermore, both these works proposed a straightforward extension of a real-valued extreme learning machine classifier, with a complex-valued kernel. Their approach assumes that each one of the complex-valued features is independent. This is clearly not the case in many electromagnetic wave imaging applications, as is showcased with THz datasets where although amplitude and phase are independently retrieved, the features in one dataset (amplitude part) are correlated with features in the other (phase part) due to a link that can be traced to the Kramers-Kronig relationship.

The proposed approach is not plagued by Fabry-Perot etalon effects (multiple reflections at air-sample interface) which would normally have to be de-embedded algorithmically prior to placing transmittance or complex insertion loss data at the input of the classifier as they would erroneously affect the phase delay values associated with the Kramers-Kronig expressions. Other advantages of the proposed approach is the reduction of errors due to the diffractive spreading of the THz beam (which is very focused in THz transient spectroscopy to improve signal to noise ratio but has as a consequence the undesirable result of making the common assumption made in optical spectroscopy of normal incidence of an angular spectrum of plane waves at the sample less valid) [25]. The approach also reduces classification artefacts due to pseudocoherence errors (constructive and destructive interference effects due to linear thickness variation in the sample across the beam aperture which has a tendency to over-estimate the sample absorbance) [26]. The proposed methodology thus complements other conventional feature-factorization classification approaches, and allows capturing interrelations among features. The THz datasets are ideal to showcase this issue.

The paper is organised as follows. In Section 1, we overview the main mathematical background regarding

complex-valued RKHS, taking into consideration recently developed Wirtinger's calculus concepts for complex RKHS. Section 2 provides a general overview of real ELM algorithms. In Sections 3, we develop the generalised complex-valued ELM and extend the simple 2D case to the multiclass classification tasks to deal with 2D and 3D inputs or even higher dimensional inputs with complex entries. This analysis forms the main contributions of the paper. Section 4 discusses a THz spectra classification task using three classification approaches: CELM, ELM, and SVMs, for validation and comparison purposes. Section 5 provides some remarks and a discussion of the proposed algorithms and section 6 provides a conclusion and directions for future research.

1. Review of complex-valued RKHS and Wirtinger's Calculus

Kernel based methods have been used as popular tools to solve non-linear classification problems in machine learning, among which Reproducing Kernel Hilbert Spaces (RKHS) algorithms play a central role. Recently, a novel class of complex-valued RKHS (CRKHS) algorithms was introduced aimed at mapping the inputs to the primal and dual Hilbert space [27]. The general class of CRKHS algorithms develop kernel algorithms in the widely linear sense [27–29] that are suitable for the processing of complex-valued data, i.e. higher dimensional datasets which are often associated with separable features in both space and time. Throughout the paper, we denote the sets of all integers, real and complex numbers by \mathcal{N} , \mathcal{R} , and \mathcal{C} respectively. We use \Re and \Im to denote the real and imaginary parts. The imaginary operator is denoted by \mathcal{J} . Complex-valued quantities appear in the caret symbol and matrix and vector valued quantities are labelled by boldfaced symbols.

1.1. Reproducing Kernel Hilbert Spaces

The CRKHS algorithm performs a nonlinear estimation by mapping the input space to an infinite dimensional feature space, where the output is a linear combination of the feature maps. The presented methods build upon kernel matrices of the form $(\hat{\psi}(\hat{z}_1), \dots, \hat{\psi}(\hat{z}_N))$, $N \in \mathcal{N}$, and \hat{z}_n ($n = 1, 2, \dots, N$) denotes the n -th set of complex valued input data. We first assume that the number of dimensions regarding each complex input equals to 2. The complex data $\hat{\mathbf{Z}} = \hat{z}_1, \hat{z}_2, \dots, \hat{z}_N$ are mapped to the so called feature space $\hat{\mathcal{H}}$ as such that $\hat{\psi} : \mathcal{C}^N \rightarrow \hat{\mathcal{H}} : \hat{\psi}(\hat{\mathbf{Z}}) = \hat{\kappa}(\cdot, \hat{\mathbf{Z}})$, where $\hat{\mathcal{H}}$ is also called complex-valued Hilbert spaces and is the complex-valued RKHS induced by the complex kernel function $\hat{\kappa} : \hat{\chi} \times \hat{\chi} \rightarrow \hat{\mathcal{F}}$ with $\hat{\mathbf{Z}} = \mathbf{X} + \mathcal{J}\mathbf{Y} \in \hat{\chi}$, and $\hat{\mathcal{F}}$ is a complex decision function. We call $\hat{\psi}$, the feature map of $\hat{\mathcal{H}}$. The approach allows for infinite dimensional Hilbert spaces to be treated. An important complex kernel is the complex Gaussian kernel, which is defined as:

$$\kappa_{\sigma_j, \mathcal{C}^N}^j(\hat{\mathbf{Z}}, \hat{\omega}) := \exp\left(-\frac{\sum_{k=1}^d (\hat{z}_k - \hat{\omega}_k^*)^2}{\sigma_j^2}\right), \quad (1)$$

where $\hat{\mathbf{Z}}, \hat{\omega} \in \mathcal{C}^N$, $N \in \mathcal{N}$ or infinite, and $\hat{\omega}$ labels complex weight (margin), with the symbol $*$ denoting a Hermitian matrix, \hat{z}_k denoting the k -th component of the complex vector $\hat{\mathbf{Z}} \in \mathcal{C}^N$ and $\exp(\cdot)$ denotes the extended exponential function in the complex domain. Here, κ^j indicates the j -th kernel function, depending on the value of kernel parameter σ_j , which in our case is varied due to the different input requirements for the normal machine learning procedure, and therefore time consuming. A proposed method is to fix the amount of kernel parameter σ_j to σ for all kernels in order to simplify computation [30].

The inner product of a complex kernel function $\hat{\kappa}$ shows its reproducing property that satisfies the following equations:

$$\hat{\kappa}(\hat{\mathbf{Z}}, \hat{\mathbf{Z}}') = (\langle \hat{\kappa}(\cdot, \hat{\mathbf{Z}}), \hat{\kappa}(\cdot, \hat{\mathbf{Z}}') \rangle_{\hat{\mathcal{H}}})^* \quad (2)$$

This equation shows that the inner product of the complex kernel function $\hat{\kappa}$ is Hermitian, indicated by $*$. We define the spaces $\mathcal{X}^2 \equiv \mathcal{X} \times \mathcal{X} \subseteq \mathcal{R}^{2n}$ and $\hat{\mathbf{Z}}, \hat{\mathbf{Z}}' \in \hat{\chi} \subseteq \mathcal{C}^n$ (two complex inputs to be classified on the basis of $\hat{\mathbf{Z}}$ and $\hat{\mathbf{Z}}'$). Then every $\hat{f} \in \hat{\mathcal{H}}$ can be viewed as a function defined on either \mathcal{X}^2 or $\hat{\chi}$ such that

$$\hat{f}(\hat{\mathbf{Z}}) = \hat{f}(\mathbf{X} + \mathcal{J}\mathbf{Y}) = \hat{f}(\mathbf{X}, \mathbf{Y}). \quad (3)$$

According to [12], the complexity of Hilbert space $\hat{\mathcal{H}} = \hat{\mathcal{F}}^{\Re} + \mathcal{J}\hat{\mathcal{F}}^{\Im}$; $\hat{\mathcal{F}}^{\Re}, \hat{\mathcal{F}}^{\Im} \in \mathcal{H}$ is a doubled real space \mathcal{H} , and satisfies $\hat{\mathcal{H}} = \mathcal{H} \times \mathcal{H} = \mathcal{H}^2$. The complexified space $\hat{\mathcal{H}}$ is a complex RKHS with complex kernel $\hat{\kappa}$, which can be represented by its respective real kernel with corresponding imaginary part of zero. The complex feature map $\hat{\psi}(\hat{\mathbf{Z}})$ of the sampled data from the complex input space to the complexified RKHS $\hat{\mathcal{H}}$ follows the equation:

$$\hat{\psi}(\hat{\mathbf{Z}}) = \hat{\psi}(\mathbf{X} + \mathcal{J}\mathbf{Y}) = \hat{\psi}(\mathbf{X}, \mathbf{Y}) = \psi(\mathbf{X}, \mathbf{Y}) + \mathcal{J}\psi(\mathbf{X}, \mathbf{Y}) \quad (4)$$

where $\psi \in \mathcal{R}$ is the feature map of the real reproducing kernel $\kappa \in \mathcal{R}$ and satisfying $\psi(\mathbf{X}, \mathbf{Y}) = \kappa(\cdot, (\mathbf{X}, \mathbf{Y}))$.

We deduce the following equation:

$$\begin{aligned} \langle \hat{\psi}(\hat{\mathbf{Z}}), \hat{\psi}(\hat{\mathbf{Z}}') \rangle_{\hat{\mathcal{H}}} &= 2\langle \psi(\mathbf{X}, \mathbf{Y}), \psi(\mathbf{X}', \mathbf{Y}') \rangle_{\mathcal{H}} \\ &= 2\kappa((\mathbf{X}', \mathbf{Y}')', (\mathbf{X}, \mathbf{Y})) \end{aligned} \quad (5)$$

We extend the two-dimensional inner complex inner product to three dimensions using the following expression:

$$\begin{aligned} \langle \hat{\psi}(\hat{\mathbf{Z}}), \hat{\psi}(\hat{\mathbf{Z}}') \rangle_{\hat{\mathcal{H}}} &= \langle \psi(\mathbf{X}, \mathbf{Y}, \mathbf{V}), \psi(\mathbf{X}', \mathbf{Y}', \mathbf{V}') \rangle_{\mathcal{H}^3} \\ &= 2\langle [\mathbf{e}'_1, \mathbf{e}'_2], [\mathbf{e}_1, \mathbf{e}_2] \rangle + 2\langle [\mathbf{e}'_2, \mathbf{e}'_3], [\mathbf{e}_2, \mathbf{e}_3] \rangle + 2\langle [\mathbf{e}'_1, \mathbf{e}'_3], [\mathbf{e}_1, \mathbf{e}_3] \rangle \end{aligned} \quad (6)$$

where $[\mathbf{e}_1, \mathbf{e}_2] = \psi(\mathbf{X}, \mathbf{Y}) = \psi(\mathbf{Y}, \mathbf{X}) = [\mathbf{X}, \mathbf{Y}]$ with $\mathbf{X}, \mathbf{Y} \in \mathcal{R}$, and we adopt similar expressions for $[\mathbf{e}_1, \mathbf{e}_3]$, $[\mathbf{e}_2, \mathbf{e}_3]$, $[\mathbf{e}'_1, \mathbf{e}'_2]$, $[\mathbf{e}'_1, \mathbf{e}'_3]$, and $[\mathbf{e}'_2, \mathbf{e}'_3]$. The bivectors \mathbf{e}_v and

\mathbf{e}'_v , $v = 1, 2, 3$ are the orthonormal basis (planes) of the two complex inputs $\hat{\mathbf{Z}}$ and $\hat{\mathbf{Z}}' = \mathcal{R}\mathbf{X}' + \mathcal{J}\mathbf{Y}' + \mathcal{K}\mathbf{V}'$, respectively, where the quantities \mathcal{R} , \mathcal{J} and \mathcal{K} are the scalar parts of a hypercomplex-like inner product structure; the kernel function with the equation $\kappa(\cdot, \cdot) = \langle \cdot, \cdot \rangle$ reproduces a Hilbert space \mathcal{H} . From Eqn. 6, we can easily deduce that the dual of the 3D complexified ELM task is equivalent to six real ELM tasks employing the kernel 2κ .

2. Real-valued Extreme Learning Machine Classifier

A real-valued extreme learning machine classifier (simplified as extreme learning machine) is a generalised single-hidden-layer feedforward network (SLFN). It maps the training data \mathbf{X} from input space to feature space using a feature map $\psi(\mathbf{X})$ with output weights ω . In contrast to SVMs, ELM theory [2, 3] shows that the item bias b should not be given in the ELM learning. Therefore, in this feature space, a linear decision function f_L is constructed as $f_L(\mathbf{X}) = \psi(\mathbf{X})\omega^v$ where $\omega^v = [\omega_1^v, \dots, \omega_L^v]^T$ is the vector of the output weights between the hidden layer of L nodes and the output node and $\psi(\mathbf{X}) = [\varphi_1(\mathbf{X}), \dots, \varphi_L(\mathbf{X})]$ is the output feature vector of the hidden layer with respect to the input \mathbf{X} . The latter maps the data from the λ dimensional input space to the L -dimensional hidden-layer feature space or ELM feature space \mathcal{H} .

Given a set of training data (x_i, ϑ_i^v) , and $\theta = [\vartheta_1^v, \dots, \vartheta_N^v]^T$ with ϑ_i^v ($i \in 1, \dots, N$) real-valued known label matrix corresponding to m classes, the function of the ELM algorithm is to minimize the training error $\|\psi\omega - \theta\|^2$ and the norm of the output weights $\|\omega\|$. The hidden-layer feature mapping matrix ψ is represented as:

$$\psi = \begin{bmatrix} \varphi_1(x_1) & \cdots & \varphi_L(x_1) \\ \vdots & & \vdots \\ \varphi_1(x_N) & \cdots & \varphi_L(x_N) \end{bmatrix} \quad (7)$$

where the size of ψ is only decided by the numbers of training samples N and the number of hidden nodes L , which is irrelevant to the number of output nodes (number of classes), and $x_1, \dots, x_N \in \mathbf{X}$.

For an m class classifier with m output nodes where $m > 1$, the classification problem (denoted by h_P) using the ELM classifier can be formulated as

$$\min : h_P = \frac{1}{2}\|\omega\|^2 + C\frac{1}{2}\sum_{i=1}^N \|\xi_i^v\|^2 \quad (8)$$

where $\xi_i^{vT} = \vartheta_i^{vT} - \psi(x_i)\omega$, $i = 1, \dots, N$, $\vartheta_i^v = \vartheta_{i,1}, \dots, \vartheta_{i,N}$, $\omega = [\omega_1^v, \dots, \omega_m^v]$, and the training error vector $\xi_i^v = [\xi_{i,1}, \dots, \xi_{i,m}]^T$. Symbol C is a penalty variable, which is a user-specific parameter and provides a trade-off between the distance of the separating margin and the training errors. When the S^{th} element $\vartheta_{i,S}$ is one and the remaining of ϑ_i^v are zero, that means the original class label is S .

The Karush-Khun-Tucker (KKT) theorem is used to solve the dual optimization problem (denoted by h_D) in order to train the ELM classifier:

$$h_D = \frac{1}{2}\|\omega\|^2 + C\frac{1}{2}\sum_{i=1}^N \|\xi_i^v\|^2 - \sum_{i=1}^N \sum_{\ell=1}^m a_{i,\ell}(\psi(x_i)\omega_\ell - \vartheta_{i,\ell} - \xi_{i,\ell}) \quad (9)$$

where ω_ℓ^v is the vector of the weights linking hidden layer to the ℓ th output node, and $\omega = [\omega_1^v, \dots, \omega_m^v]$. The Lagrange multiplier $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N]^T$ is used with each element $\mathbf{a}_i = [a_{i,1}, \dots, a_{i,m}]$ to be a vector. The rules of Wirtinger's Calculus are employed to compute the respective gradients:

$$\frac{\partial h_D}{\partial \omega_\ell^v} = 0 \rightarrow \omega = \psi^T \mathbf{A} \quad (10)$$

$$\frac{\partial h_D}{\partial \xi_i^v} = 0 \rightarrow \mathbf{a}_i = C\xi_i^v, \quad i = 1, \dots, N \quad (11)$$

$$\frac{\partial h_D}{\partial \mathbf{a}_i} = 0 \rightarrow \psi(x_i)\omega - \vartheta_i^{vT} + \xi_i^{vT} = 0, \quad i = 1, \dots, N \quad (12)$$

According to the aforementioned equations, for the case where the number of training samples is not too large ($L \gg N$) the output function of the ELM classifier is:

$$f(\mathbf{X}) = \psi(\mathbf{X})\omega = \psi(\mathbf{X})\psi^T \left(\frac{\mathbf{I}}{C} + \psi\psi^T \right)^{-1} \theta \quad (13)$$

where $\mathbf{I} = [1, 1, \dots, 1]^T$.

With $f_\ell(\mathbf{X})$ denoting the output function of the ℓ th output node and $f(\mathbf{X}) = [f_1(\mathbf{X}), \dots, f_m(\mathbf{X})]$, the predicted class label of sample \mathbf{X} is

$$\text{label}(\mathbf{X}) = \arg_{i \in 1, \dots, m} \max f_i(\mathbf{X}). \quad (14)$$

In this case where the number of training samples is very large ($N \gg L$), the output function of the ELM classifier is given from:

$$f(\mathbf{X}) = \psi(\mathbf{X})\omega = \psi(\mathbf{X}) \left(\frac{\mathbf{I}}{C} + \psi^T \psi \right)^{-1} \psi^T \theta \quad (15)$$

3. Complex-valued Extreme Machine Learning

3.1. A quaternary classification problem

An important step of machine learning is to find hyperplanes that separate the space in relation to different target classes. Using a binary classification for example, as shown in Fig. 1 a hyper-plane can be defined for Extreme Learning Machine to separate the space \mathcal{H} into two parts, \mathcal{H}_+ and \mathcal{H}_- .

Recall that in any real Hilbert space \mathcal{H} , a hyperplane may be defined, containing all the elements $f \in \mathcal{H}$ that satisfy

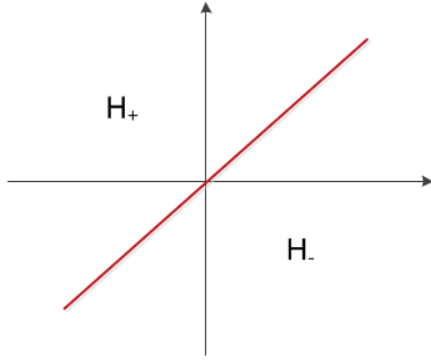


Fig. 1. Illustration of a hyperplane in EML to separate the space \mathcal{H} into two parts, \mathcal{H}_+ and \mathcal{H}_- .

$$\langle \mathbf{f}, \boldsymbol{\omega} \rangle_{\mathcal{H}} = 0, \quad (16)$$

for some $\boldsymbol{\omega} \in \mathcal{H}$. In contrast to real-valued SVM algorithms, in the real-valued ELM algorithms, the offset of the hyperplane from the origin has been removed. As shown in Fig. 1, where ELM is used to obtain binary classification of inputs, any hyperplane of \mathcal{H} divides the space into two parts, $\mathcal{H}_+ = \mathbf{f} \in \mathcal{H}; \langle \mathbf{f}, \boldsymbol{\omega} \rangle_{\mathcal{H}} > 0$ and $\mathcal{H}_- = \mathbf{f} \in \mathcal{H}; \langle \mathbf{f}, \boldsymbol{\omega} \rangle_{\mathcal{H}} < 0$. The goal of the ELM classifiers is to separate distinct classes of data by minimising the norm of output weights $\|\boldsymbol{\omega}\|$ that actually leads to a maximum margin hyper-plane [30]. In order to be able to generalize the ELM classifier to operate in a complex space, we adopt the method proposed in [12], and define a complex hyper-plane that divides the complex space $\hat{\mathcal{H}}$ into four parts by introducing a Hermitian matrix, label *. This approach enables the classification of objects into four classes (instead of two), as shown in Fig. 2. In imaging applications, spatial features in adjacent pixels can be retained, i.e we have real and imaginary components in x and y directions. According to [12], we have $\langle \hat{\mathbf{f}}, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} = \langle \hat{\mathbf{f}}^{\Re}, \hat{\boldsymbol{\omega}}^{\Re} \rangle_{\mathcal{H}} + \langle \hat{\mathbf{f}}^{\Im}, \hat{\boldsymbol{\omega}}^{\Im} \rangle_{\mathcal{H}} + \mathcal{I}(\langle \hat{\mathbf{f}}^{\Im}, \hat{\boldsymbol{\omega}}^{\Re} \rangle_{\mathcal{H}} - \langle \hat{\mathbf{f}}^{\Re}, \hat{\boldsymbol{\omega}}^{\Im} \rangle_{\mathcal{H}})$ where $\hat{\mathcal{H}} = \mathcal{H}^2$, $\hat{\mathbf{f}}, \hat{\boldsymbol{\omega}}$ indicates a complex decision function and the corresponding margin of the hyper-plane $\hat{\mathcal{H}}$. The symbol of $\langle \cdot \rangle_{\mathcal{H}}$ denotes the inner product in the corresponding real-valued input space. It can be represented by its respective real kernel with corresponding imaginary part of zero. It satisfies the equation:

$$\begin{aligned} \kappa(\mathbf{X}, \mathbf{Y}) &= \langle \kappa(\cdot, \mathbf{X}), \kappa(\cdot, \mathbf{Y}) \rangle_{\mathcal{H}} \\ &= \kappa(\langle \cdot, \mathbf{Y} \rangle, \kappa(\cdot, \mathbf{X}))_{\mathcal{H}} = \kappa^*(\mathbf{X}, \mathbf{Y}) \end{aligned} \quad (17)$$

The Hermitian matrix, indicated by *, is a matrix with complex entries and is equal to its own conjugate transpose. The kernel $\kappa(\cdot, \mathbf{X})$ is used for a feature map of real-valued input space \mathcal{H} , labelled by $\psi(\mathbf{X})$. The corresponding complex kernel $\hat{\kappa}(\cdot, \hat{\mathbf{Z}})$ is a linear combination of real value kernels. We use symbols $\hat{\psi}(\hat{\mathbf{Z}})$ to denote complex feature mapping in the context. In this as well as in the following section, in addition to the bold-faced symbols used for the vector and matrix valued quantities, we use the symbol caret above all the complex-valued quantities, which are related to ma-

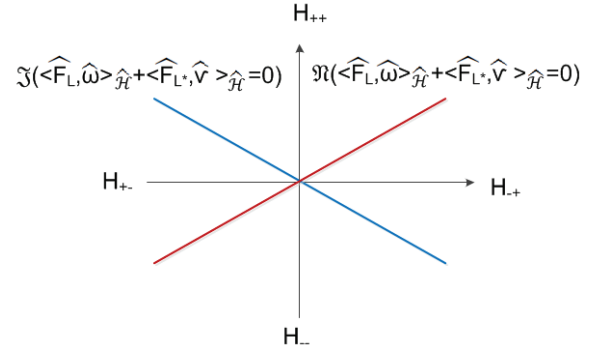


Fig. 2. Illustration of a complex couple of hyper-planes in ELM to divide the complex space $\hat{\mathcal{H}}$ into four parts by introducing Hermitian matrix, label *. Symbols of $\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\nu}}$ denote the output weights in relation to decision function of $\hat{\mathbf{f}}_L$ and its couple $\hat{\mathbf{f}}_{L^*}$.

trices, with specific subscripts to denote the corresponding row and/or column of the complex-valued entries.

Definition 1. Let $\hat{\mathcal{H}}$ be a complex Hilbert space. The complex couple of hyperplanes is defined as the set of all $\mathbf{f} \in \hat{\mathcal{H}}$ that satisfy one of the following relations

$$\Re(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) = 0 \quad (18a)$$

$$\Im(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) = 0 \quad (18b)$$

for some $\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\nu}} \in \hat{\mathcal{H}}$, where $\hat{\mathbf{f}}_L \in \hat{\mathcal{H}}$ represents two hyperplanes of the doubled real space, \mathcal{H}^2 .

The input space is divided into four portions which are associated with the complex couple of hyper-planes as shown below:

$$\mathcal{H}_{++} = \begin{cases} \Re(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) > 0 \\ \Im(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) > 0 \end{cases} \quad (19a)$$

$$\mathcal{H}_{+-} = \begin{cases} \Re(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) > 0 \\ \Im(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) < 0 \end{cases} \quad (19b)$$

$$\mathcal{H}_{-+} = \begin{cases} \Re(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) < 0 \\ \Im(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) > 0 \end{cases} \quad (19c)$$

$$\mathcal{H}_{--} = \begin{cases} \Re(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) < 0 \\ \Im(\langle \hat{\mathbf{f}}_L, \hat{\boldsymbol{\omega}} \rangle_{\hat{\mathcal{H}}} + \langle \hat{\mathbf{f}}_{L^*}, \hat{\boldsymbol{\nu}} \rangle_{\hat{\mathcal{H}}}) < 0 \end{cases} \quad (19d)$$

We now extend the formulation to 3D complex inputs. The 3D complex inputs are orthogonally projected into three 2D complex input spaces along horizontal (\mathbf{e}_1), vertical (\mathbf{e}_2) and frontal (\mathbf{e}_3) directions. As a result, we obtain three decision functions: $\hat{\mathbf{f}}_{Le_1}, \hat{\mathbf{f}}_{Le_2}, \hat{\mathbf{f}}_{Le_3}$ and the associated output weights $\hat{\boldsymbol{\omega}}_{e_1}, \hat{\boldsymbol{\omega}}_{e_2}, \hat{\boldsymbol{\omega}}_{e_3}$. The three decision functions span the subspace of the three real hyperplanes $\mathcal{H}_{\mathbf{E}}$, $\mathbf{E} = \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. That is, $\hat{\mathbf{f}}_{Le_1} \in \mathcal{H}_{e_1}$, $\hat{\mathbf{f}}_{Le_2} \in \mathcal{H}_{e_2}$, $\hat{\mathbf{f}}_{Le_3} \in \mathcal{H}_{e_3}$. As the three coupled hyperplanes form 3D hyperspaces,

in the following equations, we use Ω instead of \mathcal{H} to represent the real hyperspace in three dimensions. The complexified hyperspace $\hat{\omega}$ in three dimensions is equivalent to triple \mathcal{H} . The coupled parts consist of $\hat{f}_{Le1}^*, \hat{f}_{Le2}^*, \hat{f}_{Le3}^*$ and $\hat{\omega}_{e1}, \hat{\omega}_{e2}, \hat{\omega}_{e3}$. As shown in Fig. 3, we label the coordinate planes of $x-y$, $x-z$ and $y-z$ with $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, respectively. We use r, j, k instead of $-, +$ to label the 3D complex hyperspaces $\hat{\omega}$ split to eight quadrants. The symbols of r, j, k are associated with x -axis, y -axis, z -axis, respectively, and satisfy the functions: $r = \mathbf{e}_2\mathbf{e}_3$, $j = \mathbf{e}_3\mathbf{e}_1$, and $k = \mathbf{e}_1\mathbf{e}_2$. These complex hyper-spaces are represented as follows:

$$\Omega_{rrj} = \begin{cases} \mathcal{R}(\langle \hat{f}_{Le1}, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le1}^*, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}}) > 0 \\ \mathcal{R}(\langle \hat{f}_{Le2}, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le2}^*, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e2}}) > 0 \\ \mathcal{J}(\langle \hat{f}_{Le3}, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le3}^*, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e3}}) > 0 \end{cases} \quad (20a)$$

$$\Omega_{jrk} = \begin{cases} \mathcal{J}(\langle \hat{f}_{Le1}, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le1}^*, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}}) > 0 \\ \mathcal{R}(\langle \hat{f}_{Le2}, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le2}^*, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e2}}) > 0 \\ \mathcal{K}(\langle \hat{f}_{Le3}, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le3}^*, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e3}}) > 0 \end{cases} \quad (20b)$$

$$\Omega_{rkk} = \begin{cases} \mathcal{R}(\langle \hat{f}_{Le1}, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le1}^*, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}}) < 0 \\ \mathcal{K}(\langle \hat{f}_{Le2}, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le2}^*, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e2}}) > 0 \\ \mathcal{K}(\langle \hat{f}_{Le3}, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le3}^*, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e3}}) > 0 \end{cases} \quad (20c)$$

$$\Omega_{jkj} = \begin{cases} \mathcal{J}(\langle \hat{f}_{Le1}, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le1}^*, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}}) < 0 \\ \mathcal{K}(\langle \hat{f}_{Le2}, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le2}^*, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e2}}) > 0 \\ \mathcal{J}(\langle \hat{f}_{Le3}, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le3}^*, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e3}}) > 0 \end{cases} \quad (20d)$$

$$\Omega_{-rkk} = \begin{cases} \mathcal{R}(\langle \hat{f}_{Le1}, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le1}^*, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}}) > 0 \\ \mathcal{K}(\langle \hat{f}_{Le2}, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le2}^*, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e2}}) < 0 \\ \mathcal{K}(\langle \hat{f}_{Le3}, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le3}^*, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e3}}) < 0 \end{cases} \quad (20e)$$

$$\Omega_{-jkj} = \begin{cases} \mathcal{J}(\langle \hat{f}_{Le1}, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le1}^*, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}}) > 0 \\ \mathcal{K}(\langle \hat{f}_{Le2}, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le2}^*, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e2}}) < 0 \\ \mathcal{J}(\langle \hat{f}_{Le3}, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le3}^*, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e3}}) < 0 \end{cases} \quad (20f)$$

$$\Omega_{-rrj} = \begin{cases} \mathcal{R}(\langle \hat{f}_{Le1}, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le1}^*, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}}) < 0 \\ \mathcal{R}(\langle \hat{f}_{Le2}, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le2}^*, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e2}}) < 0 \\ \mathcal{J}(\langle \hat{f}_{Le3}, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le3}^*, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e3}}) < 0 \end{cases} \quad (20g)$$

$$\Omega_{-jrk} = \begin{cases} \mathcal{J}(\langle \hat{f}_{Le1}, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le1}^*, \hat{\omega}_{e1} \rangle_{\mathcal{H}_{e1}}) < 0 \\ \mathcal{R}(\langle \hat{f}_{Le2}, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le2}^*, \hat{\omega}_{e2} \rangle_{\mathcal{H}_{e2}}) < 0 \\ \mathcal{K}(\langle \hat{f}_{Le3}, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e1}} + \langle \hat{f}_{Le3}^*, \hat{\omega}_{e3} \rangle_{\mathcal{H}_{e3}}) < 0 \end{cases} \quad (20h)$$

Here, \mathcal{H}_- denotes the opposite site associated with the complex hyper-plane surfaces \mathcal{H} . The decision functions

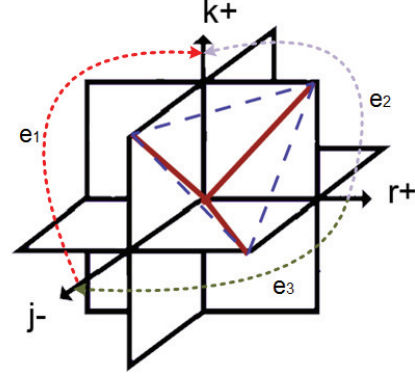


Fig. 3. Illustration of a rational complex hyperspace consisting of three complex couples of hyperplanes calculated via ELMs.

$\hat{f}, \hat{g}, \hat{d}$ are implemented using the same kernel with different coordinates. As shown in Fig. 3, the three complex couples (indicated by three red solid lines) of hyper-planes in the fourth quadrant form a rational complex hyper-plane (a 3D rational complex hyper-space). The complex hyperspaces divide the 3D complex entries into eight parts. For classification of $m > 1$ classes, these inputs will be further grouped with total $2^{3-1} \times m$ classes via ELM.

We formulate the complex ELM classification task as follows. The given complex-valued training samples consist of eight separate classes $\Lambda_{rrj}, \Lambda_{jrk}, \Lambda_{rkk}, \Lambda_{jkj}, \Lambda_{-rkk}, \Lambda_{-jkj}, \Lambda_{-rrj}, \Lambda_{-jrk}$ such as $\{(\hat{z}_n, \vartheta_n); n = 1, \dots, N\} \subset \Psi \times \{\pm r \pm j \pm k\}$. That is: if $\vartheta_n = +\mathcal{R} + \mathcal{R} + \mathcal{J}$, then $\hat{z}_n \in \Lambda_{rrj}$; If $\vartheta_n = +\mathcal{J} + \mathcal{R} + \mathcal{K}$, then $\hat{z}_n \in \Lambda_{jrk}$; If $\vartheta_n = -\mathcal{R} + \mathcal{K} + \mathcal{K}$, then $\hat{z}_n \in \Lambda_{rkk}$; If $\vartheta_n = -\mathcal{J} + \mathcal{K} + \mathcal{J}$, then $\hat{z}_n \in \Lambda_{jkj}$; If $\vartheta_n = +\mathcal{R} - \mathcal{K} - \mathcal{K}$, then $\hat{z}_n \in \Lambda_{-rkk}$; If $\vartheta_n = +\mathcal{J} - \mathcal{K} - \mathcal{J}$, then $\hat{z}_n \in \Lambda_{-jkj}$; If $\vartheta_n = -\mathcal{R} - \mathcal{R} - \mathcal{J}$, then $\hat{z}_n \in \Lambda_{-rrj}$; If $\vartheta_n = -\mathcal{J} - \mathcal{R} - \mathcal{K}$, then $\hat{z}_n \in \Lambda_{-jrk}$. Here, $\hat{z}_n = \mathbf{X} + \mathcal{J}\mathbf{Y} + \mathcal{K}\mathbf{Z}$, $\Lambda^- + \Lambda = \mathbf{0}$, and $\Lambda, \Psi \in \hat{\omega}$.

3.2. Determination of the maximum-margin hyper-planes of CELM

The goal of the complex machine learning task is to estimate a complex couple of maximum margin hyperplanes. According to the work [12], for a 2D simple case, it is to minimise

$$\left\| \hat{\omega}^r + \hat{\nu}^r \right\|_{\mathcal{H}^2}^2 + \left\| -(\hat{\omega}^j + \hat{\nu}^j) \right\|_{\mathcal{H}^2}^2 = 2(\|\hat{\omega}\|_{\mathcal{H}}^2 + \|\hat{\nu}\|_{\mathcal{H}}^2).$$

In a similar manner to the ELM algorithm, for the classification task, CELM is used to minimize the training error $\hat{\delta} = \|\psi\hat{\omega} + \psi^*\hat{\nu} - \hat{\theta}\|^2$ with $\hat{\theta} = [\hat{\vartheta}_1, \dots, \hat{\vartheta}_N]^T$ being a known complex-valued label matrix corresponding to m classes, as well as the norm of the output margins (weights). Therefore, the primal complex EML optimization problem can be formulated as

$$\min_{(\hat{\omega}, \hat{\nu}, C)} = \left(\frac{1}{2} \|\hat{\omega}\|_{\mathcal{H}}^2 + \frac{1}{2} \|\hat{\nu}\|_{\mathcal{H}}^2 + \frac{C}{N} \sum_{n=1}^N (\hat{\delta}_n^2) \right), \quad (21)$$

Subject to:

$$\begin{cases} \Re(\langle \hat{\psi}_{\mathcal{H}}(\hat{z}_n), \hat{\omega} \rangle + \langle \hat{\psi}_{\mathcal{H}}^*(\hat{z}_n), \hat{\nu} \rangle) \geq \hat{\vartheta}_n^r - \hat{\delta}_n^r \\ \Im(\langle \hat{\psi}_{\mathcal{H}}(\hat{z}_n), \hat{\omega} \rangle + \langle \hat{\psi}_{\mathcal{H}}^*(\hat{z}_n), \hat{\nu} \rangle) \geq \hat{\vartheta}_n^j - \hat{\delta}_n^j \end{cases} \quad (22)$$

where C is a parameter given by users aiming for a trade-off between the distance in relation to the separating margin and the training error and $n \in 1, \dots, N$. Using positive Lagrangian multipliers \mathbf{a} and \mathbf{b} , the Lagrangian function becomes:

$$\begin{aligned} L(\hat{\omega}, \hat{\nu}, \mathbf{a}, \mathbf{b}) = & \frac{1}{2} \|\hat{\omega}\|_{\mathcal{H}}^2 + \frac{1}{2} \|\hat{\nu}\|_{\mathcal{H}}^2 + \frac{C}{N} \|\hat{\delta}_{n,\rho}^r + \hat{\delta}_{n,\rho}^j\|^2 \\ & - \sum_{n=1}^N \sum_{\rho=1}^m a_{n,\rho} \\ & \left(\Re(\langle \hat{\psi}_{\mathcal{H}}(\hat{z}_n), \hat{\omega}_{\rho} \rangle + \langle \hat{\psi}_{\mathcal{H}}^*(\hat{z}_n), \hat{\nu}_{\rho} \rangle) - \hat{\vartheta}_{n,\rho}^r + \hat{\delta}_{n,\rho}^r \right) \\ & - \sum_{n=1}^N \sum_{\rho=1}^m b_{n,\rho} \\ & \left(\Im(\langle \hat{\psi}_{\mathcal{H}}(\hat{z}_n), \hat{\omega}_{\rho} \rangle + \langle \hat{\psi}_{\mathcal{H}}^*(\hat{z}_n), \hat{\nu}_{\rho} \rangle) - \hat{\vartheta}_{n,\rho}^j + \hat{\delta}_{n,\rho}^j \right) \end{aligned} \quad (23)$$

Here, $\hat{\theta}_n = \{\hat{\vartheta}_{n,\rho}\}$ with $\rho = 1, \dots, m$ and $n \in 1, \dots, N$, where $\{\hat{\vartheta}_{n,\rho}\}$ denotes the output value of the ρ^{th} output node for the training data \hat{z}_n and m labels the number of the classes of the output. When both the real and imaginary parts of S^{th} element $\hat{\vartheta}_{n,S}$ are one and the remaining elements of $\hat{\vartheta}_n$ are zero, that means the original class label is $S + \mathcal{J}S$. According to Wirtinger's Calculus to compute the respective gradients, we have

$$\begin{aligned} \frac{\partial L}{\partial \hat{\omega}_{\rho}^*} = & \frac{1}{2} \hat{\omega}_{\rho} - \frac{1}{2} \sum_{n=1}^N a_{n,\rho} \hat{\psi}_{\mathcal{H}}^T(\hat{z}_n) + \frac{\mathcal{J}}{2} \sum_{n=1}^N b_{n,\rho} \hat{\psi}_{\mathcal{H}}^T(\hat{z}_n) = 0 \\ \Rightarrow & \hat{\omega}_{\rho} = \sum_{n=1}^N (\alpha_{n,\rho} - \mathcal{J}b_{n,\rho}) \hat{\psi}_{\mathcal{H}}^T(\hat{z}_n) \end{aligned} \quad (24)$$

$$\begin{aligned} \frac{\partial L}{\partial \hat{\nu}_{\rho}^*} = & \frac{1}{2} \hat{\nu}_{\rho} - \frac{1}{2} \sum_{n=1}^N a_{n,\rho} \hat{\psi}_{\mathcal{H}}^{*T}(\hat{z}_n) + \frac{\mathcal{J}}{2} \sum_{n=1}^N b_{n,\rho} \hat{\psi}_{\mathcal{H}}^{*T}(\hat{z}_n) = 0 \\ \Rightarrow & \hat{\nu}_{\rho} = \sum_{n=1}^N (a_{n,\rho} - \mathcal{J}b_{n,\rho}) \hat{\psi}_{\mathcal{H}}^{*T}(\hat{z}_n) \end{aligned} \quad (25)$$

$$\begin{aligned} \frac{\partial L}{\partial \hat{\delta}_{n,\rho}^r} = & \frac{2C}{N} \hat{\delta}_{n,\rho}^r - a_{n,\rho} = 0 \Rightarrow \hat{\delta}_{n,\rho}^r = \frac{N}{2C} a_{n,\rho} \\ \frac{\partial L}{\partial \hat{\delta}_{n,\rho}^j} = & \frac{2C}{N} \hat{\delta}_{n,\rho}^j - b_{n,\rho} = 0 \Rightarrow \hat{\delta}_{n,\rho}^j = \frac{N}{2C} b_{n,\rho} \end{aligned} \quad (26)$$

$$\begin{aligned} \frac{\partial L}{\partial a_{n,\rho}} = & -\frac{1}{2} \left(\Re(\langle \hat{\psi}_{\mathcal{H}}(\hat{z}_n), \hat{\omega}_{\rho} \rangle + \langle \hat{\psi}_{\mathcal{H}}^*(\hat{z}_n), \hat{\nu}_{\rho} \rangle) \right) \\ & - \frac{1}{2} \left(-\hat{\vartheta}_{n,\rho}^r + \hat{\delta}_{n,\rho}^r \right) = 0 \end{aligned} \quad (27)$$

$$\begin{aligned} \frac{\partial L}{\partial b_{n,\rho}} = & -\frac{1}{2} \left(\Im(\langle \hat{\psi}_{\mathcal{H}}(\hat{z}_n), \hat{\omega}_{\rho} \rangle + \langle \hat{\psi}_{\mathcal{H}}^*(\hat{z}_n), \hat{\nu}_{\rho} \rangle) \right) \\ & - \frac{1}{2} \left(-\hat{\vartheta}_{n,\rho}^j + \hat{\delta}_{n,\rho}^j \right) = 0 \end{aligned} \quad (28)$$

According to the last two equations,

$$\langle \hat{\psi}_{\mathcal{H}}(\hat{z}_n), \hat{\omega}_{\rho} \rangle + \langle \hat{\psi}_{\mathcal{H}}^*(\hat{z}_n), \hat{\nu}_{\rho} \rangle - \hat{\vartheta}_{n,\rho}^r + \hat{\delta}_{n,\rho}^r = 0 \quad (29)$$

By substituting Eqn. 24, Eqn. 25 and Eqn. 26, Eqn. 29 can be written as

$$\begin{aligned} (\mathbf{a} - \mathcal{J}\mathbf{b}) \left(\hat{\psi}_{\mathcal{H}}(\hat{z}_n) \hat{\psi}_{\mathcal{H}}^T(\hat{z}_n) + \hat{\psi}_{\mathcal{H}}^*(\hat{z}_n) \hat{\psi}_{\mathcal{H}}^{*T}(\hat{z}_n) \right) \\ + \frac{N}{2C} (\mathbf{a} + \mathcal{J}\mathbf{b}) = \hat{\theta} \end{aligned} \quad (30)$$

The output decision functions of the CELM classifier are:

$$\begin{aligned} \Re(\hat{f}_L(\hat{\mathbf{Z}})) = & \Re(\langle \hat{\psi}_{\mathcal{H}}(\hat{\mathbf{Z}}), \hat{\omega} \rangle) \\ = & \Re \left(\hat{\psi}_{\mathcal{H}}(\hat{\mathbf{Z}}) \hat{\psi}_{\mathcal{H}}^T \left(\left(\hat{\psi}_{\mathcal{H}} \hat{\psi}_{\mathcal{H}}^T + \hat{\psi}_{\mathcal{H}}^* \hat{\psi}_{\mathcal{H}}^{*T} \right) + \frac{N}{2C} \mathbf{I} \right)^{-1} \hat{\theta} \right) \end{aligned} \quad (31a)$$

$$\begin{aligned} \Im(\hat{f}_L(\hat{\mathbf{Z}})) = & \Im(\langle \hat{\psi}_{\mathcal{H}}(\hat{\mathbf{Z}}), \hat{\omega} \rangle) \\ = & -\Im \left(\hat{\psi}_{\mathcal{H}}(\hat{\mathbf{Z}}) \hat{\psi}_{\mathcal{H}}^T \left(\left(\hat{\psi}_{\mathcal{H}} \hat{\psi}_{\mathcal{H}}^T + \hat{\psi}_{\mathcal{H}}^* \hat{\psi}_{\mathcal{H}}^{*T} \right) + \frac{N}{2C} \mathbf{I} \right)^{-1} \hat{\theta} \right) \end{aligned} \quad (31b)$$

$$\begin{aligned} \Re(\hat{f}_{L^*}(\hat{\mathbf{Z}})) = & \Re(\langle \hat{\psi}_{\mathcal{H}}^*(\hat{\mathbf{Z}}), \hat{\nu} \rangle) \\ = & \Re \left(\hat{\psi}_{\mathcal{H}}^*(\hat{\mathbf{Z}}) \hat{\psi}_{\mathcal{H}}^{*T} \left(\left(\hat{\psi}_{\mathcal{H}} \hat{\psi}_{\mathcal{H}}^T + \hat{\psi}_{\mathcal{H}}^* \hat{\psi}_{\mathcal{H}}^{*T} \right) + \frac{N}{2C} \mathbf{I} \right)^{-1} \hat{\theta} \right) \end{aligned} \quad (31c)$$

$$\begin{aligned} \Im(\hat{f}_{L^*}(\hat{\mathbf{Z}})) = & \Im(\langle \hat{\psi}_{\mathcal{H}}^*(\hat{\mathbf{Z}}), \hat{\nu} \rangle) \\ = & -\Im \left(\hat{\psi}_{\mathcal{H}}^*(\hat{\mathbf{Z}}) \hat{\psi}_{\mathcal{H}}^{*T} \left(\left(\hat{\psi}_{\mathcal{H}} \hat{\psi}_{\mathcal{H}}^T + \hat{\psi}_{\mathcal{H}}^* \hat{\psi}_{\mathcal{H}}^{*T} \right) + \frac{N}{2C} \mathbf{I} \right)^{-1} \hat{\theta} \right) \end{aligned} \quad (31d)$$

where $\hat{\psi}_{\mathcal{H}} \hat{\psi}_{\mathcal{H}}^T$ and $\hat{\psi}_{\mathcal{H}}^* \hat{\psi}_{\mathcal{H}}^{*T}$ are $N \times N$ or $L \times L$ matrices, according to the size of the inputs. The predicted class label of sample $\hat{\mathbf{Z}}$:

$$\text{label}(\hat{\mathbf{Z}}) = \arg \max_{n=1, \dots, m} (\langle \hat{\psi}_{\mathcal{H}}(\hat{\mathbf{Z}}), \hat{\omega} \rangle + \langle \hat{\psi}_{\mathcal{H}}^*(\hat{\mathbf{Z}}), \hat{\nu} \rangle) \quad (32)$$

where $\text{label}(\hat{\mathbf{Z}}) = \text{label}(\Re(\hat{\mathbf{Z}})) + \mathcal{J}\text{label}(\Im(\hat{\mathbf{Z}}))$. Here, we employ the induced real kernel $2\hat{\kappa}^r$ instead of the complex kernel $\hat{\kappa}$ for the solution of the complex labeling function.

For the 3D multiclass classification, the complex labelling function can be written as the linear combination of 2D multiclass classification tasks along three orthogonal coordinate planes. It satisfies the following expression:

$$\begin{aligned} \text{label}(\hat{\mathbf{Z}}) = \arg \max_{n=1, \dots, m} & \langle \hat{\psi}(\hat{\mathbf{Z}}), \hat{\omega}_{e_1} \rangle_{\hat{\mathcal{H}}_{e_1}} + \langle \hat{\psi}^*(\hat{\mathbf{Z}}), \hat{\omega}_{e_1} \rangle_{\hat{\mathcal{H}}_{e_1}} \\ & + \langle \hat{\psi}(\hat{\mathbf{Z}}), \hat{\omega}_{e_2} \rangle_{\hat{\mathcal{H}}_{e_2}} + \langle \hat{\psi}^*(\hat{\mathbf{Z}}), \hat{\omega}_{e_2} \rangle_{\hat{\mathcal{H}}_{e_2}} \\ & + \langle \hat{\psi}(\hat{\mathbf{Z}}), \hat{\omega}_{e_3} \rangle_{\hat{\mathcal{H}}_{e_3}} + \langle \hat{\psi}^*(\hat{\mathbf{Z}}), \hat{\omega}_{e_3} \rangle_{\hat{\mathcal{H}}_{e_3}} \end{aligned} \quad (33)$$

This is consistent with Subsection 3.1.

The block diagram of the proposed algorithm is illustrated in Fig. 4. The complex extreme learning machine enables multi-class classification when both the input and output variables as well as the optimisation variables are all complex-valued. The label for a given testing sample from multi-class is complex-valued and has the highest output value of the complex-coupled output decision functions \hat{f}_L and \hat{f}_L^* for the given testing sample.

4. CELM classification of powdered sample spectra

In this section, we discuss the application of CELM to tackle a difficult classification problem of spectra from a range of powdered samples. These are: sand, talcum, salt, powdered sugar, wheat flour, and baking soda. A well set-up T-ray imaging system which generate femtosecond duration terahertz pulses is used to detect the T-ray sample responses [6, 31]. The transmittance of these samples is studied by broadband time-domain terahertz transient spectrometry. The measurement of samples by T-ray pulsed spectroscopy is conducted in The University of Adelaide [6], Australia. The aim of the work is to demonstrate the classification performance of the proposed algorithmic scheme, and compare it with standard real-valued classification techniques, such as SVMs and ELMs. The use of T-rays for the classification of powdered samples has been viewed as a basic emerging technology for illicit substance detection (e.g., metamphetamines, cocaine, MDMA, as well as explosives such as HMX, RDX, C-4, and TNT or bio-hazards such as anthrax spores etc) and has important security implications[32]. Furthermore, this technology is currently being assessed for its ability to detect counterfeit drugs or patent infringement rights by pharmaceutical companies worldwide (e.g. distinguishing acetamidophenol from other isomorphs may be possible). In addition, terahertz pulses penetrate well the plastic cases where drugs are encapsulated so it may be also possible to perform quick tests to establish alterations of their physical properties after storage. This is particularly useful from a quality control perspective as well as for assessing whether expired drugs may still be functional and freely distributed to the third world well beyond their expiry dates. There is a significant humanitarian motivation to assess whether such practices could be adopted. The current study paves the way towards such investigations by only focusing on the

ability of T-ray spectroscopy to detect different densities, thicknesses, and concentrations of specific powder samples. The powder recognition tasks are conducted using compressed pellets 4 mm thick for six different powdered substances. Compression of the pellets and measurements are performed using a standardised protocol that ensures repeatable compression, under controlled humidity conditions.

A 2-D T-ray image of the sample can be obtained after inserting a powder sample at the measuring port of the interferometric spectrometer and comparing its transmission to that of free space (background interferogram). Fourier transforming the sample as well as background interferograms is then performed and the resulting spectra are ratioed to calculate the complex insertion loss. Care is taken to only use spectral images where the sample fills entirely the aperture of the imaging system. Signal processing is then applied to assign the key features of training vectors for different classes of signals. Since the Fourier transformation produces complex-valued spectra, containing both phase and magnitude information, our aim is to retain such information in the classification process. This approach elucidates further one of the advantages of performing spectroscopic investigations at the THz part of the spectrum, that the associated wavelengths probing the sample are several micrometers and up to few millimetres long, so that phase stability is ensured throughout the experiments. This is not possible using optical or infrared measurement modalities where phase information may only be inferred by resorting to calculations based on the Kramers-Kronig relationships. A further advantage of the longer wavelengths associated with such measurements is that scattering is minimized, so that sample absorbance may be directly inferred from reflectance and transmittance measurements.

In the current investigations, the observed magnitude and phases at certain key frequencies constitute pairs of feature subsets on which a classification scheme can be designed for the separation of powder samples. An important advantage of this approach is the small dimensionality of feature vectors, allowing for the features to be directly extracted from pulsed responses using relatively low computational load. This is particularly important as future THz imaging systems are expected to be capable of recording significantly larger numbers of pixels per image. For the complex-valued classification task, we view the extracted magnitude and phases as real and imaginary parts, respectively, therefore the extracted features are complex-valued. A leave-one-out (LOO) [33, 34] estimator within the T-ray data set is utilized to provide a nearly unbiased estimate of the prediction error rate. Due to the relatively small number of measurements for the different powders, LOO is preferred as the overall classification experiment is averaged over more runs. In this experiment, accuracy of classification is used as the quantity for assessing the performance of all the classification tasks, and it equals to the number of correct classified test vectors N_L^{true} for all-class samples $\iota = 1, \dots, m$ versus the total number of vectors to be tested

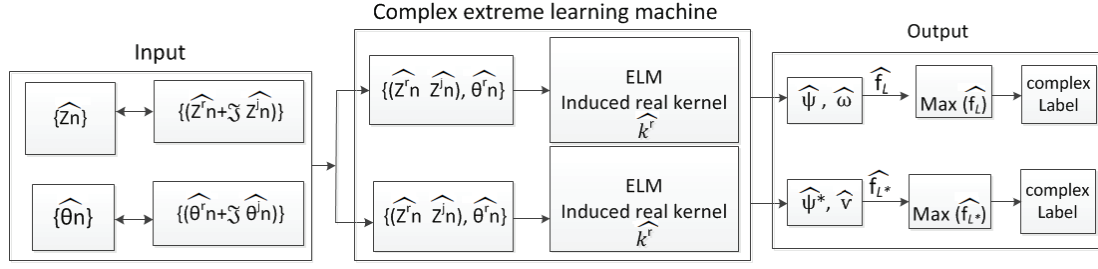


Fig. 4. Illustration of the procedure for 2D complex-valued learning via complex extreme learning machine. The system facilitates multi-class classification when both input and output variables are complex-valued. The approach also uses optimisation variables that are complex-valued. The label for a given testing sample from the multi-class is a complex-valued entity which has the highest output value associated with the complex-coupled output decision functions \hat{f}_L and \hat{f}_{L^*} for the given testing sample.

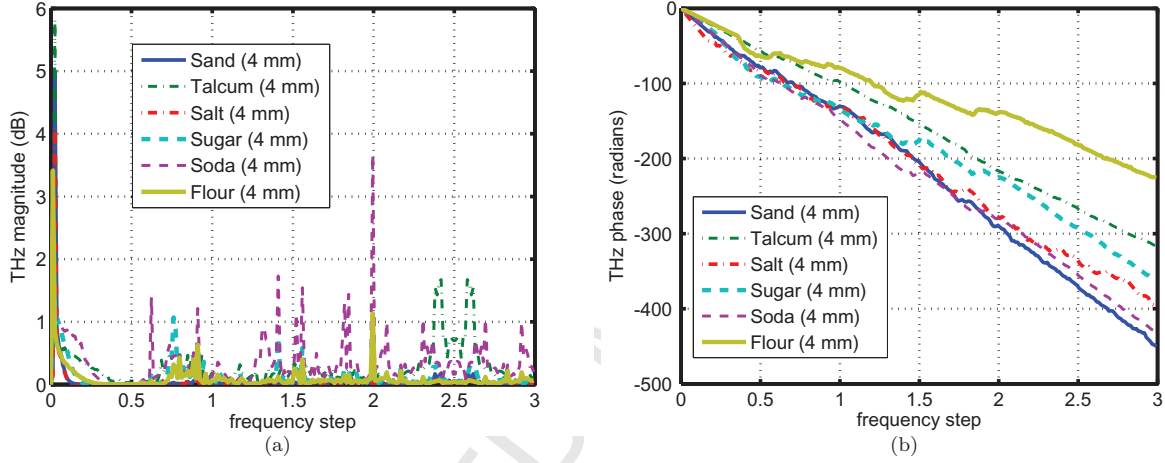


Fig. 5. Illustration of Fourier spectrum. (a) Illustrates the magnitude (attenuation) as a function of terahertz frequency. (b) Illustrates corresponding phase delay (equivalent to chromatic dispersion) as a function of terahertz frequency.

N^{total} .

$$\text{accuracy} = \frac{\sum_{i=1}^m N_i^{\text{true}}}{N^{\text{total}}} \quad (34)$$

4.1. Fourier transform based spectral analysis and multi-class classification

Each image of a powder sample consists of $6 \times 50 = 300$ pixels. For each pixel, the number of time domain samples is 400 (corresponding to a rather coarse frequency resolution after Fourier transformation). Fig. 5 shows the phase and magnitude plots in the frequency domain from one pixel of powder image datasets, with a cut-off frequency at 3 THz. Obvious phase changes associated with the six types of powder samples can be observed in Fig. 5 (b). We produce the learning vector patterns for multiclass recognition using CELM, as shown in Fig. 6 after Fourier transformation of the time-domain signatures and extraction of the corresponding complex-valued features associated with the second frequency bin. We use 49 training samples for each powder sample for training the classifier. Two real RKHS ker-

nels are used for mapping. The labels are complex-valued and produce 12 classes for powder materials. The background colour shows clear contour shape of the derived decision surface, (these are numbered from 2-12), and correspond to the magnitude calculations derived from the sum of real and imaginary values of the complex labels (argument values). Only the labels consisting of the same real and imaginary parts (both parts label the same class) are assigned to the class of power sample. The labelled contours that correspond to different real and imaginary parts (the real and imaginary parts label the different classes) are illustrated in Fig. 7. These regions are undecided in the classification process and are therefore excluded to avoid over-fitting problems.

Fig. 8 illustrates the multi-classification result for the six types of powder substances using an SVMs classifier. SVMs are designed according to a pair wise strategy. One real Gaussian kernel is used to map the input data into a 2D Fourier feature space. The support vectors indicated by cyan circles are subsets of the training data sets and are used to construct a two-dimensional hyper-plane in feature space, which acts as a boundary separating each class of

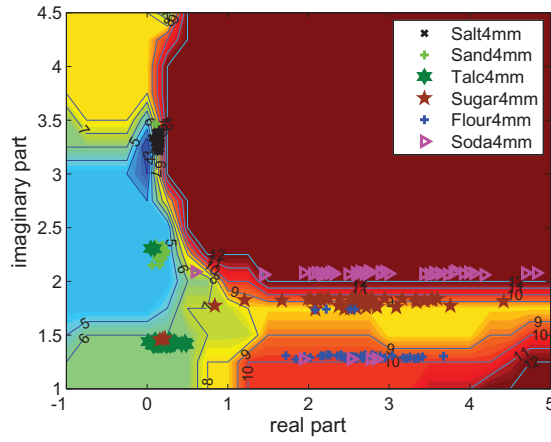


Fig. 6. Complex-valued learning vectors regarding the six-class examples measured by T-rays are plotted for the illustration of the linear decision function among each classes by applying induced real RKHS kernels to map the complex input data into 2D complex-valued feature space. There are 49 pixels selected randomly from each of six classes of powder samples. The labels are complex-valued that produce 12 classes for powder materials.

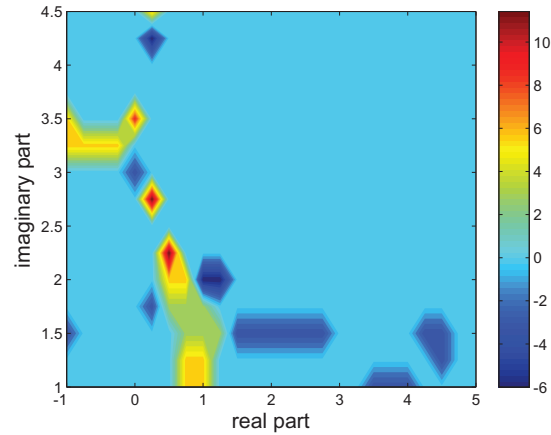


Fig. 7. Illustration of the colour coded regions with non-zeros indicated by the colour bar. Colour regions with non-zero values indicate that the multi-class powder sample classification process remains undecided by CELM as the real and imaginary parts are not equal to each other.

different powder materials.

4.2. Comparison of classification results

Tables 1–2 show the achieved multi-class classification accuracy on the THz Fourier spectral features after applying two types of machine learning algorithms: CELM and SVMs via pairwise classification, respectively, with varying penalty parameter C and elapsed time. A leave-one-out (LOO) estimator is used for both training and testing purposes. For training purposes, pixels from all the classes are presented to the two classifiers. The remaining 1 pixel

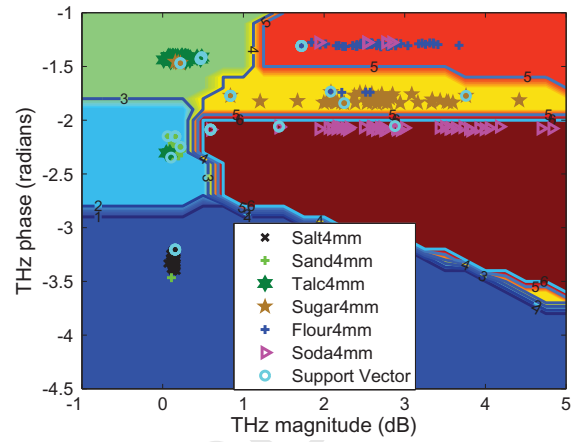


Fig. 8. Learning vectors for the six-class examples plotted to illustrate the linear decision function between the pairs of classes after applying a Gaussian kernel for mapping. There are 49 pixels selected randomly from each of six classes of powder samples. Background colour shows clearly the contour shape of the decision surface. The small yellow region on the bottom of the right hand side remains undecided.

from each class is used for testing. The classification experiments are repeated over 50 runs. Therefore, the test elapsed time indicates the 300 runs required to perform classification as testing time. All the powder classification runs are performed using MATLAB version R2013a on a personal computer running Windows 7 with an Intel(R) Core(TM) i5-3470 CPU (3.20 GHz) and 8 GB of memory.

We use a real-valued Gaussian kernel for all classifiers, with parameter σ of 100 which showed through several classification attempts that it produced the best results after adopting a trial-and-error approach [35]. By setting this parameter value to 100, we illustrate comparable classification performance among different classifiers. The penalty parameter C is varied from 10^6 to 10 for CELM in steps of 1 on a log scale, excluding $C=100$ after adopting the above-mentioned approach. For SVMs, the classification accuracy is not varied with parameter C when $\sigma = 100$. The classification performance for CELM is listed in Table. 1, where both real and imaginary parts are used as input features for training and testing the classifier. The known real and imaginary parts of complex-valued classification labels within an associated matrix are applied further to calculate the real and imaginary parts of the output weights, respectively, for training purposes. For real-valued machine learning, i.e. SVMs, we use both phase and magnitude as training and testing feature vectors, to find the maximum possible classification accuracy, with only real values as an input to the classification label matrix. This way, one can make a useful comparison with the classification accuracy using CELMs, as these use both real and imaginary portions of the input complex-valued labels each time for training, separately. The elapsed time of the CELM classification scheme is e^{-1} seconds, which is nearly one hundred times faster than real SVMs classification which re-

quires an elapsed time of over 30 seconds. Furthermore, the classification accuracy of CELM is 99.7%, which is much better than the classification accuracy of SVMs (88.1%) when the number of support vectors is around 1500. In order to improve the classification accuracy of SVMs, we set the parameter σ to 0.001. The classification performance regarding SVMs is listed in Table 2. For the measurement of classification performance regarding SVMs, we also list the number of support vectors used to calculate the boundaries of each of the powder classes. It can be observed that when the classification experiment becomes difficult to be conducted with the decreased value of penalty parameter C , the elapsed time associated with SVMs is much longer due to the increased computation load from the increased number of support vectors. Generally, the classification accuracy of CELM is still comparable with SVMs, and the elapsed time of CELM is 10 times faster than SVMs. Moreover, the classification performance of CELMs can be further improved by further adjusting the Gaussian kernel parameter. Overall, a significantly superior classification performance is achieved using CELM as opposed to SVMs.

In order to compare the classifier performances between complex-valued CELM and real-valued ELM, we extend the same classifier design from CELM to real-valued ELM. Since the output needs to be associated to a real-valued parameter, only part of the complex-valued inputs are used in the training process of the classifier, (either the real or the imaginary part). In the current case, we illustrate the resultant real-valued classification performance with use of the imaginary part, (phase), as the feature vector, to train and test the classifier. This is deemed acceptable as both the real and imaginary parts in our experiments are related to each other through the Kramers-Kronig relations. We use phase as the input feature space, as shown in Fig. 5(b). The classification performance is listed in Table. 3. One can observe an improved classification accuracy with the increased value of the C variable. The maximum classification accuracy is found to be 93.2% at $C = 10^{12}$. The elapsed time of e^{-2} second is 10 times faster than that required when using CELM, however, the classification accuracy using real-valued ELM is lower than using complex-valued ELM which achieves a maximum classification accuracy of 99.7%.

The classification performance is listed in Table 3. It can be seen that the observed variation in classification accuracy may be attributed to the penalty parameter C . It can also be seen that the classification accuracy using real-valued ELM is lower than that achieved using complex-valued ELM.

5. Discussion and general remarks

The following remarks are made regarding the proposed classification procedure and elucidate the contributions of this paper to the classifiers literature. Firstly, as described in Eqn. 24, the dimension L of the output weight $\hat{\omega}$, $\hat{\nu}$ in

the CELM that link the hidden layer is not changed and not decided by the output $\hat{\theta}$. Secondly, similar to ELM classifiers, the feature map $\hat{\psi}(\hat{\mathbf{Z}})$ is randomly generated, and the size of kernel used to generate CELM is equal to N , the size of the entire data set in all cases. Thirdly, in CELM, the size of kernel matrix $\hat{\psi}\hat{\psi}^T = N \times N$ and its dual $\hat{\psi}^* \hat{\psi}^{*T} = N \times N$ is fixed, when the number of the hidden nodes L is much smaller than the number of training samples: $L \ll N$. For the case when $L \gg N$, as would be the case for large scale applications with multiple input classes, the kernel matrix and its dual can be written as $\hat{\psi}^{*T} \hat{\psi}^* = L \times L$. This algorithmic feature results in low computation costs. Furthermore, the rational multiclass classification of high-dimensional, i.e. \mathcal{N} input data is equivalent to $2 \times \mathcal{N}$ real ELM classification tasks after employing \mathcal{N} 2D real kernel 2κ , according to Eqn. 6. Using three-dimensional inputs as an example, CELM is achieved after conducting six real ELM classification procedures along three orthogonal coordinate planes \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 . Another aspect of our approach is that we have formulated a complex couple hyperspace that separates the complex volume space from 2D to 3D through the cascaded application of real-valued ELM classifiers. The output value and output weights according to 2D inputs are formulated using induced real-valued kernels. We have also shown that it is straightforward to extend the two-dimensional output values to 3D in order to perform multiclass classification, with the complex labelling function being viewed as the linear combination of 2D multiclass classification along three orthogonal coordinate planes.

Finally, in our simulations depicting the classification of 6 powder samples we used complex-valued labels to produce 12 classes of spectral signatures on the basis of T-ray spectroscopic measurements. The elapsed time of CELM classification scheme reaches e^{-1} seconds, which is nearly hundred times faster than real SVMs classification. Furthermore, the classification accuracy of CELM is comparable with that of SVMs. This is a significant result. In addition, due to the complex-valued labels produced via CELM, only the labels consisting of the same real and imaginary parts are validated for final pattern identification. This avoids over-fitting problems. In the THz classification study, we illustrate these regions with labels consisting of different real and imaginary parts as regions undecided by CELM. This approach enables us to visualise regions where CELM can be inconclusive. When compared with a real ELM classification scheme, CELM shows increased classification accuracy due to the higher dimension of the input space.

6. Conclusion

The purpose of a complex ELM classifier is to solve complex-valued multi-classification tasks via complexified coupled RKHS. This paper investigated the rational ELM with extension from the 2D complexified coupled RKHS \mathcal{H} to a high dimensional space to directly solve multi-class classification problems. The proposed approach allows us

Table 1

Classification accuracy (%) and elapsed time are illustrated for powder classification using CELM with varying penalty parameter C . The Fourier spectral features are extracted as real and imaginary parts.

| C | each class name and classification accuracy | | | | | | Total classification accuracy | elapsed time (s) |
|---------|---|------|--------|-------|-------|------|-------------------------------|------------------|
| | salt | sand | talcum | sugar | flour | soda | | |
| 1000000 | 100 | 100 | 100 | 98 | 100 | 100 | 99.7 | 0.6398 |
| 100000 | 100 | 100 | 100 | 73 | 100 | 100 | 95.5 | 0.6373 |
| 10000 | 100 | 100 | 100 | 0 | 100 | 100 | 82.6 | 0.6385 |
| 1000 | 100 | 100 | 100 | 0 | 100 | 95.8 | 81.9 | 0.6479 |
| 10 | 100 | 0 | 100 | 0 | 97.9 | 79 | 61 | 0.6371 |

Table 2

Classification accuracy (%) and elapsed time are illustrated for powder classification using SVMs with varying penalty parameter C . The Fourier spectral features are extracted as input features for classification.

| C | each class name and classification accuracy | | | | | | Total classification accuracy | # of SVs | elapsed time (s) |
|-----|---|------|--------|-------|-------|-------|-------------------------------|----------|------------------|
| | salt | sand | talcum | sugar | flour | soda | | | |
| 10 | 100 | 100 | 100 | 98 | 100 | 100 | 99.7 | 101 | 7.1292 |
| 5 | 100 | 100 | 100 | 98 | 100 | 48.98 | 98 | 143 | 7.9249 |
| 1 | 100 | 100 | 100 | 91.8 | 100 | 48.98 | 90.1 | 320 | 11.2009 |
| 0.5 | 100 | 100 | 100 | 95.9 | 100 | 40.82 | 89.4 | 473 | 14.4457 |

Table 3

Classification accuracy (%) and elapsed time for powder classification using real ELM with varying penalty parameter C . The same validation procedure to that used in CELM validation is followed. One of the Fourier spectral features, magnitude, is extracted as training and testing feature vectors for all classes of powder datasets.

| C | each class name and classification accuracy | | | | | | Total classification accuracy | elapsed time (s) |
|-----------|---|------|--------|-------|-------|------|-------------------------------|------------------|
| | salt | sand | talcum | sugar | flour | soda | | |
| 10^8 | 100 | 0 | 12.2 | 100 | 100 | 100 | 68.7 | 0.0066 |
| 10^9 | 91.8 | 0 | 67.3 | 100 | 100 | 100 | 76.5 | 0.0066 |
| 10^{10} | 91.8 | 0 | 100 | 100 | 100 | 100 | 82.3 | 0.0067 |
| 10^{11} | 100 | 32.6 | 100 | 100 | 100 | 100 | 88.8 | 0.0064 |
| 10^{12} | 100 | 59.2 | 100 | 100 | 100 | 100 | 93.2 | 0.0066 |
| 10^{13} | 98 | 55.1 | 100 | 87.8 | 100 | 100 | 90.1 | 0.0066 |
| 10^{14} | 10.2 | 20 | 38.8 | 49 | 55 | 47 | 36.7 | 0.0068 |

to explore the advantages of using a more complex structure of a high-dimensional input space by solving several cascaded real ELM classification tasks. In this sense, the work can be seen as an extension of other methodologies for multi-objective machine learning [36–38] that complement recent advances in feature extraction literature [1, 39–41]

The experimental work that is conducted within the proposed context has illustrated that using the CELM classification scheme to solve a multi-class classification problem of THz spectral measurements of 6 powdered samples, the elapsed time reached e^{-1} seconds, this is nearly a hundred times faster than that achieved using real-valued SVMs classification, without sacrificing classification accuracy. Furthermore, the current paper offers a systematic way for studying complex kernel matrix classifiers in higher dimensions. Complex-valued ELM processing data according to both the magnitude and the orientation leads to complex-valued labels, and allows us to perform classifica-

tion tasks of multi-class samples by assigning for our samples twice the original number of classes in the classifier input. The current method shows improved classification accuracy compared with a real-valued ELM classification scheme. This is aligned to our expectations since a richer information input space is presented to the classifier. The work also presents a systematic method to design complex labels for the classification of different classes of powder samples while avoiding over-fitting problems. This should prove particularly valuable to classification problems frequently encountered in the THz literature, and has also important applications in areas as diverse as chemometrics that would benefit from real-time implementation [42, 42–46].

Acknowledgments

The authors would like to thank B. Ferguson of the ARC National T-ray Facility, University of Adelaide, for providing the THz spectra.

References

- [1] J. M. Leiva-Murillo and A. Artes-Rodríguez, "Information-theoretic linear feature extraction based on kernel density estimators: A review," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **42**(6), pp. 1180–1189, 2012.
- [2] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **42**(2), pp. 513–529, 2011.
- [3] G.-B. Huang, D. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics* **2**(2), pp. 107–122, 2011.
- [4] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery* **2**, pp. 121–167, 1998.
- [5] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, USA, 1995.
- [6] X. Yin, B. Ng, B. Fischer, B. Ferguson, and D. Abbott, "Support vector machine applications in terahertz pulsed signals feature sets," *IEEE Sensors Journal* **7**(12), pp. 1597–1608, 2007.
- [7] X. Yin, B. Ng, and D. Abbott, *Terahertz imaging for biomedical applications: pattern recognition and tomographic reconstruction*, Springer-Verlag, New York, USA, 2012.
- [8] Y. Iye, "Composite fermions and bosons: An invitation to electron masquerade in Quantum Hall," *Proceedings of the National Academy of Sciences of the United States of America* **96**, pp. 8821–8822, 1999.
- [9] K. R. Waters, J. Mobley, and J. G. Miller, "Causality-imposed (Kramers-Kronig) relationships between attenuation and dispersion," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **52**(5), pp. 822–823, 2005.
- [10] A. El-Gindy and G. Hadad, "Nonparametric bayes error estimation using unclassified samples," *Journal of AOAC International* **95**(3), pp. 609–623, 2012.
- [11] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing* **70**(1-3), pp. 489–501, 2006.
- [12] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive learning in complex reproducing kernel hilbert spaces employing wirtinger's subgradients," *IEEE Transactions on Neural Networks and Learning Systems* **2**(99), pp. 260–276, 2012.
- [13] G. Scheuermann, H. Krüger, M. Menzel, and A. Rockwood, "Visualizing nonlinear vector field topology," *IEEE Transactions on Visualization and Computer Graphics* **4**(2), pp. 109–116, 1998.
- [14] H. Li, *Invariant Algebras and Geometric Reasoning*, World Scientific, Singapore, 2008.
- [15] T. Nitta, *Complex-Valued Neural Networks: Utilizing High-Dimensional Parameters*, Information Science Reference, Pennsylvania, 2009.
- [16] E. Hitzler, T. Nitta, and Y. Kuroe, "Applications of clifford's geometric algebra," *Advances in Applied Clifford Algebras* **23**(2), pp. 377–404, 2013.
- [17] V. Dietrich, K. Habetha, and G. Jank, *Clifford Algebras and Their Application in Mathematical Physics*, Kluwer Academic Publishers, Springer, Dordrecht, The Netherlands, 1998.
- [18] E. Corrochano and G. Sobczyk, *Geometric Algebra With Applications in Science and Engineering*, Birkhäuser Boston, c/o Springer-Verlag, New York, 2001.
- [19] B. Sethuraman, "Division algebras and wireless communication," *Notices of the American Mathematical Society* **57**(11), pp. 1432–1439, 2010.
- [20] A. Hirose, *Complex-Valued Neural Networks*, World Scientific Publishing, Singapore, 2003.
- [21] E. Bayro-Corrochano, "Geometric neural computing," *IEEE Transactions on Neural Networks* **12**(5), pp. 968–986, 2001.
- [22] E. Bayro-Corrochano and N. Arana-Daniel, "Clifford support vector machines for classification, regression, and recurrence," *IEEE Transaction on Neural Networks* **21**(11), pp. 1731–1746, 2010.
- [23] M.-B. Li, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "Fully complex extreme learning machine," *Neurocomputing* **68**, pp. 306–314, 2005.
- [24] R. V. Babu and S. Suresh, "Fully complex-valued ELM classifiers for human action," *The 2011 International Joint Conference on Neural Networks*, pp. 2803–2808, 2011.
- [25] G. C. W. J. W. Bowen and S. Hadjiloucas, "Sample-induced beam distortions in terahertz time domain spectroscopy and imaging systems," *Joint 32nd International Conference on Infrared and Millimetre Waves and 15th International Conference on Terahertz Electronics*, pp. 208–209, 2007.
- [26] S. Hadjiloucas, L. Karatzas, and J. Bowen, "Measurements of leaf water content using terahertz radiation," *IEEE Transactions on Microwave Theory and Techniques* **47**(2), pp. 142–149, 1999.
- [27] F. A. Tobar, A. Kuh, and D. P. Mandic, "A novel augmented complex valued kernel LMS," in *the 7th IEEE Sensor Array and Multichannel Workshop 2012*, pp. 473–476, 2012.
- [28] A. Shilton and D. Lai, "Quaternionic and complex-valued support vector regression for equalization and function approximation," *International Joint Conference on Neural Networks (IJCNN)*, pp. 920 – 925, 2007.
- [29] P. Bouboulis and S. Theodoridis, "Extension of

- Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS," *IEEE Transactions on Signal Processing* **59**(3), pp. 964–978, 2011.
- [30] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **42**(2), pp. 513–529, 2012.
- [31] X.-X. Yin, B. Ng, B. Ferguson, and D. Abbott, "Wavelet based local tomographic image using terahertz techniques," *Digital Signal Processing* **19**(4), pp. 750–763, 2009.
- [32] B. Ferguson, S. Wang, H. Zhong, D. Abbott, and X.-C. Zhang, "Powder refection with t-ray imaging," in *Proceeding of SPIE Terahertz for Military and Security Application, 2003*, **5070**, pp. 7–16, 2003.
- [33] K. Fukunaga and D. L. Kessell, "Chemometrics in pharmaceutical analysis: an introduction, review, and future perspectives," *IEEE Transactions on Information Theory* **IT-19**(4), pp. 434–440, 1973.
- [34] K. Fukunaga and D. M. Hummels, "Leave-one-out procedures for nonparametric error estimates," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(4), pp. 421–423, 1989.
- [35] R. R. Jackson, C. M. Carter, and M. S. Tarsitano, "Trial-and-error solving of a confinement problem by a jumping spider, portia fimbriata," *Behaviour* **138**(10), pp. 1215–1234, 2001.
- [36] Y. J. Y. Jin and B. Sendhoff, "Pareto-based multi-objective machine learning: An overview and case studies," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **38**(3), pp. 397–415, 2008.
- [37] J. D. Rodríguez, A. Perez, D. Arteta, D. Tejedor, and J. A. Lozano, "Using multidimensional bayesian network classifiers to assist the treatment of multiple sclerosis," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **42**(6), pp. 1705–1715, 2012.
- [38] H. Q. Vu, G. Li, N. S. Sukhorukova, G. Beliaikov, S. Liu, C. Philippe, H. Amiel, and A. Ugon, "K-complex detection using a hybrid-synergic machine learning method," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **39**(4), pp. 1478–1490, 2012.
- [39] D. Muni, N. Pal, and J. Das, "Genetic programming for simultaneous feature selection and classifier design," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **36**(1), pp. 106–117, 2006.
- [40] K. Mao, "Identifying critical variables of principal components for unsupervised feature selection," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **35**(2), pp. 339–344, 2005.
- [41] S. Salcedo-Sanz, G. Camps-Valls, F. Perez-Cruz, J. Sepulveda-Sanchis, and C. Bousono-Calzon, "Enhancing genetic feature selection through restricted search and walsh analysis," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **34**(4), pp. 398–406, 2004.
- [42] J. Park, S. Baek, S. Bae, and M. Jeong, "Dual features functional support vector machines for fault detection of rechargeable batteries," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **39**(4), pp. 480–485, 2009.
- [43] O. Omitaomu, M. Jeong, A. Badiru, and J. Hines, "Online support vector regression approach for the monitoring of motor shaft misalignment and feed-water low rate," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **37**(5), pp. 962–970, 2007.
- [44] P. Maji and S. Paul, "Rough sets for selection of molecular descriptors to predict biological activity of molecules," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **40**(6), pp. 639–648, 2010.
- [45] G. Duan and Y.-W. Chen, "A machine learning-based framework for automatic visual inspection of micro-drill bits in PCB production," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **42**(6), pp. 1679–1689, 2012.
- [46] Y.-S. Jeong, I.-H. Kang, M.-K. Jeong, and D. Kong, "A new feature selection method for one-class classification problems," *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications & Reviews* **39**(4), pp. 1500–1509, 2012.

short authors' biographies : X.X. Yin, S Hadjiloucas, Y Zhang, J. He, Y. Wang and David Zhang

X.X. Yin: was born in Dalian, China. She is now a research fellow in Victoria University, Australia. She received the B.Eng. degree in industrial electronics from Dalian University, China, and the Ph.D. degree from the University of Adelaide, Adelaide, Australia, in 2009, on three-dimensional terahertz computed tomography under D. Abbott and B. Ng. She was a Visiting Scholar at the University of Reading, U.K., under S. Hadjiloucas and at the University of Cambridge, U.K., under L. F. Gladden in 2008. Dr. Yin received a series of awards, including the Ph.D. Scholarship of the University of Adelaide (2005), the WavE Travel Fellowship from Switzerland (2006), the International Association for Pattern Recognition (IAPR) travel stipend from Hong Kong (2006), the Research Abroad Scholarship from the University of Adelaide (2006), the Roger Pysden Memorial Fellowship from Australia Business, ABL State Chamber (2006), a student scholarship from 2007 the IEEE International Symposium on Industrial Electronics, Vigo, Spain (2007), a Mutual Community Postgraduate Travel Grant from the University of Adelaide (2007), an Overseas Travel Fellowship from the Australian Research Council Nanotechnology Network (ARCNN) (2007), and the Gertrude Rohan Prize (2010). In 2009, she won a prestigious Australian Research Council (ARC) postdoctoral fellowship.

S Hadjiloucas: was appointed Lecturer in Cybernetics in January 2000. His appointment follows two previous appointments as an EC TMR (Training and Mobility of Researchers) Post-Doctoral Research Fellow working on Terahertz Instrumentation as part of the INTERACT (International Terahertz Action) project in UK and Germany. His work aims to enhance the existing equipment infrastructure at Reading and develop novel on-line and off-line signal processing modalities. Work so far has concentrated in developing a range of sensing techniques across the electromagnetic spectrum and a range of novel signal processing algorithms. Most of the work has important applications in the biomedical field. Prior to completion of his PhD (1996) at the Department of Cybernetics, he obtained BSc and MPhil degrees from the Department of Pure and Applied Biology at the University of Leeds. He has considerable experience in cross-disciplinary research across all natural Sciences.

Y Zhang: is the Director of the Centre for Applied Informatics (CAI), and coordinates a multidisciplinary e-research program across the University. He is an international research leader in databases, data mining, health informatics, web information systems, and web services. He has published over 220 research papers in international journals and conferences proceedings, and authored/edited 12 books.

His research has been supported by a number of Australian Research Council (ARC) linkage projects and discovery project grants. His research has made significant impacts on society. For example, the multidisciplinary research into e-health has produced software systems and mapping tools to assist relevant government/industry organisations establish health needs, allowing the development of policy based on firm evidence.

J. He: is currently an Associate Professor with tenure in the College of Engineering and Science, Victoria University. She has been awarded a PhD degree from Academy of Mathematics and System Science, Chinese Academy of Sciences in 2006. Prior to joining to Victoria University, she worked in University of Chinese Academy of Sciences, China during 2005-2008. She has been active in areas of Data Mining, Web service/Web search, Spatial and Temporal Database, Multiple Criteria Decision

Making, Intelligent System, Scientific Workflow and some industry field such as E-Health, Petroleum Exploration and Development, Water resource Management and e-Research. She has published over 60 research papers in refereed international journals and conference proceedings including ACM transaction on Internet Technology (TOIT), IEEE Transaction on Knowledge and Data Engineering (TKDE), Information System, The Computer Journal, Computers and Mathematics with Applications, Concurrency and Computation: Practice and Experience, International Journal of Information Technology & Decision Making, Applied Soft Computing, and Water Resource Management. She received research fund from ARC DECRA, ARC Discovery, ARC Linkage, Youth Science Fund of Chinese Academy of Sciences in 2007, National Science Foundation of China in 2006 and etc. She served on three program committees of international conferences: International Conference on Computational Science (ICCS), The IEEE International Conference on Data Mining (ICDM), and International Symposium on Knowledge and Systems Science (KSS), as well as the program chair of HIS 2011, 2012 and 2013 conference and workshop co-chair on Wise 2012, HIS 2012, APWeb 2008, WI 2009, MCDM 2009 and etc. In addition, she has been serving as external reviewers for several international journals and conferences, such as Management Science, The Computer Journal, IEEE Transaction on Systems, Man, Cybernetics, International Journal of Information Technology and Decision Making, Journal of Management Review (in Chinese), Decision Support System, Science (in China), WISE, ICCS, ICDM, KSS, APWeb, PAKDD etc.

Y. Wang: received a BA in Nuclear Physics from Fudan University, Shanghai in 1986, and a PhD in Physics from the University of Wisconsin at Madison in 1994. After a postdoc and promotion to Assistant Professor at Mayo Clinic, he joined the faculty of Cornell University Medical College in 1997. Currently he holds the Faculty Distinguished Professorship at the Department of Radiology, and is a tenured Professor of Physics in Radiology, Professor of Biomedical Engineering, and the Principal Investigator of the 3T MRI facility at Cornell University.

David Zhang: graduated in Computer Science from Peking University. He received his MSc in 1982 and his PhD in 1985 in both Computer Science from the Harbin Institute of Technology (HIT), respectively. From 1986 to 1988 he was a Postdoctoral Fellow at Tsinghua University and then an Associate Professor at the Academia Sinica, Beijing. In 1994 he received his second PhD in Electrical and Computer Engineering from the University of Waterloo, Ontario, Canada. Currently, he is a Chair Professor at the Hong Kong Polytechnic University where he is the Founding Director of the Biometrics Technology Centre (UGC/CRC) supported by the Hong Kong SAR Government in 1998. He also serves as Visiting Chair Professor in Tsinghua University and HIT, and Adjunct Professor in Shanghai Jiao Tong University, Peking University, National University of Defense Technology and the University of Waterloo. He is the Founder and Editor-in-Chief, International Journal of Image and Graphics (IJIG); Book Editor, Springer International Series on Biometrics (KISB); Organizer, the first International Conference on Biometrics Authentication (ICBA); Associate Editor of more than ten international journals including IEEE Transactions; Technical Committee Chair of IEEE SMC and the author of more than 10 books and 300 international journal papers. Recently, he is selected in the list of Highly Cited Researchers 2014 (Total 3,200 in 21 fields of the sciences & Social Sciences). Professor Zhang is a Croucher Senior Research Fellow, Distinguished Speaker of the IEEE Computer Society, and a Fellow of both IEEE and IAPR.