Recognition of Pharmacological Bi-Heterocyclic Compounds by Using Terahertz Time Domain Spectroscopy and Chemometrics

Maciej Roman Nowak [1],* [ID]**, Rafał Zdunek [1]** [ID]**, Edward Pliński [1]** [ID]**, Piotr Świątek [2]** [ID]**, Małgorzata Strzelecka [2]** [ID]**, Wiesław Malinka [2]** [ID] **and Stanisława Plińska [3]**

[1] Faculty of Electronics, Wroclaw University of Science and Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
[2] Department of Chemistry of Drugs, Faculty of Pharmacy, Wroclaw Medical University, Wroclaw Ludwika Pasteura 1, 50-367 Wroclaw, Poland
[3] Department of Inorganic Chemistry, Faculty of Pharmacy, Wroclaw Medical University, Wroclaw Ludwika Pasteura 1, 50-367 Wroclaw, Poland
* Correspondence: maciej.nowak@pwr.edu.pl; Tel.: +48-71-320-23-66

check for updates

Abstract: In this study, we presented the concept and implementation of a fully functional system for the recognition of bi-heterocyclic compounds. We have conducted research into the application of machine learning methods to correctly recognize compounds based on THz spectra, and we have described the process of selecting optimal parameters for the kernel support vector machine (KSVM) with an additional 'unknown' class. The chemical compounds used in the study contain a target molecule, used in pharmacy to combat inflammatory states formed in living organisms. Ready-made medical products with similar properties are commonly referred to as non-steroidal anti-inflammatory drugs (NSAIDs) once authorised on the pharmaceutical market. It was crucial to clearly determine whether the tested sample is a chemical compound known to researchers or is a completely new structure which should be additionally tested using other spectrometric methods. Our approach allows us to achieve 100% accuracy of the classification of the tested chemical compounds in the time of several milliseconds counted for 30 samples of the test set. It fits perfectly into the concept of rapid recognition of bi-heterocyclic compounds without the need to analyse the percentage composition of compound components, assuming that the sample is classified in a known group. The method allows us to minimize testing costs and significant reduction of the time of analysis.

Keywords: (THz-TDS) terahertz time domain spectroscopy; (KSVM) kernel support vector machine; (kNN) k-nearest neighbour; bi-heterocyclic compounds; chemometrics

1. Introduction

A heterocyclic compound is where the atoms form a ring, with at least one non-carbon atom in a single cyclic system. Elements such as oxygen, nitrogen and sulphur appear most often in the ring [1]. Heterocyclic compounds are common in the life of every human being [2,3]. The best examples are amino acids (histamine, serotonin, melatonin, hydroxyproline), vitamins (e.g., folic acid or vitamin M, biotin = vitamin H, thiamine = vitamin B1, riboflavin = B2, niacin = B3, pyridoxine = B6, cyanocobalamin = B12), coenzymes (flavin adenine dinucleotide (FAD), reduced form FAD (FADH2), nicotinamide adenine dinucleotide phosphate (NADP+)), dyes (hem/haemoglobin, chlorophyll a), nucleic acids (adenine, guanine, cytosine, thymine, uracil) directly related to deoxyribonucleic acids (DNA) and ribonucleic acids (RNA)), insecticides (e.g., fipronil), toxins (ergotamine, divicine, muscarine) and natural and synthetic flavourings [4]. Analysing these examples,

it can be concluded that a large part of the applications find their place in medicine. It is estimated that the group of compounds containing at least one heterocyclic ring constitutes over 70% of all pharmaceuticals [5]. Heterocyclic rings are extremely important structural elements of numerous pharmaceutical preparations currently available in pharmacies and constitute the basis for searching for new biologically active substances [6]. An excellent example is pyrimidine and its derivatives used to treat Parkinson's disease, neuropathy, bacteria, viruses, fungi, cancer and hypertension [7]. Drugs based on the so-called quinazoline skeleton have similar properties. Quinazoline is a structure formed by combining two aromatic hexavalent rings: benzene and pyrimidine, which means a bi-heterocyclic structure. Some quinazoline derivatives can additionally treat inflammatory states, allergies, autoimmune diseases, eye diseases and even lower blood glucose levels [8]. A greater variety of heterocyclic structures, methods of their production and detailed applications are described in numerous specialist books related to pharmacy and medical chemistry [1,8–10].

1.1. Test Methods Used to Diagnose Drugs

The techniques used for drug diagnosis usually include Raman spectroscopy [11,12] and near infrared spectroscopy (NIR) [13,14]. Both techniques allow at-line, online and in-line measurements, which makes them very practical and useful in pharmacy and medical chemistry [15]. Liquid chromatography (LC) [16], X-ray diffraction (XRD) [17], and mass spectroscopy [18,19] are also commonly used in chemistry to examine the percentage composition of elements. Nuclear quadrupole resonance spectroscopy (NQR) plays a particularly important role [20]. The radio frequency (RF) method is used to identify counterfeit medicines in packages [21]. Drugs, especially recreational drugs, can also be diagnosed using neutron scattering techniques, in which the densities of H, C, N, O atoms in the substance are determined [22]. In the case of prohibited drugs such as methacvalone (Mandrax), cocaine, LSD, heroin, and morphine, the diagnosis is based on the detection of high concentrations of hydrogen and carbon and low concentrations of nitrogen and oxygen [23].

The last technique mentioned by us is THz spectroscopy, which is the basis of our research. When synthesizing new molecules (drug candidates), other large and complex molecules (molecular tails) are attached to the basic molecule (core). In the THz band, we observe how various tails affect the resulting spectrum. In the IR band and especially NIR, one can observe vibrations between particular atoms, which can often disturb spectral observations and even make the interpretation of the results more difficult. Only the examination of intermolecular vibrations makes it possible to distinguish individual synthesized molecules, and this is the main advantage of THz spectroscopy. The best example is the work that presents the results on the usage of THz and IR spectra for the analysis of three derivatives of the 1,4-naphthoquinone compound (plumbagin, juglone i menadione) used in the traditional Chinese medicine. The IR spectra for all three compounds were very similar and required a detailed analysis of all functional groups, while the THz range allowed for recognizing the compounds (molecular fingerprints) unambiguously. Many scientific publications have been devoted to the use of terahertz techniques for the detection and identification of prohibited drugs [24,25]. Using the Terahertz time spectroscopy (THz-TDS) system with imaging and knowledge of the location of absorption peaks of drugs, it is possible to detect prohibited psychoactive substances [26,27]. The terahertz parametric oscillator (TPO) [28] and injection-seeded terahertz parametric generation (is-TPG) methods [29] deserve special recognition. They make it possible to detect, among other things, illegal drugs and many reagents, as well as the inspection of the pharmaceutical production process with a dynamic range of about four to more than eight orders of magnitude. This allows for imaging of objects that are thicker than a thin envelope. Another application of terahertz pulsed imaging (TPI) is the study of drug coatings by means of terahertz pulse imaging (TPI) allowing for three-dimensional characterization of the tablet structure [30,31]. Coating with a thin layer of polymer is used, among others, to improve the stability of parameters, facilitate swallowing and change the drug release time [32]. This method can be successfully used to inspect the top layer thickness of tablets on the production line in real time [33,34]. Monitoring of drugs during production and storage is also possible by examining the

crystalline structure of samples with the use of terahertz radiation [35]. This is particularly important in the event of changes in environmental conditions (temperature, humidity, etc.) in the production process and may affect the final composition and action of the drug. Research on polymorphic forms of drugs and descriptions of terahertz techniques can be found in numerous publications [36,37]. The terahertz technique may help identify molecular structures of drugs with subtle differences in functional groups [38]. The authors point to the high effectiveness of the diagnosis of some diabetic drug pharmacophores in relation to density functional theory (DFT) limitations. Research was also conducted on the diagnosis of amino acids in food supplements [39]. It should be noted that terahertz techniques can be used as autonomous or in many applications they support other techniques. The use of mixed technique, a combination of Raman spectroscopy and TPI imaging spectroscopy [40] allowed the development of a method of drug release time prediction. Detailed applications of THz are described in review publications [41–44] and books [45–47]. The THz-TDS system has also been used in [48] for predicting physicochemical properties of the drugs based on piroxicam. In this work, the Partial Least Squares (PLS) -based predictive model was built to quantify the relationship between the spectra and the melting point, which can guide the selection of early drug candidates.

1.2. Drug Detection Using Terahertz Spectroscopy Combined with Support Vector Machine (Svm)

The recognition of drugs on the basis of their terahertz spectra requires the use of an appropriate mathematical apparatus. Statistical methods are useful, especially the algorithms that are based on pattern recognition methods. Numerous papers were devoted to the use of the SVM algorithm for drug analysis. The basic role of SVM is to determine whether a given structure is a pharmaceutical drug or not [49,50], and, if such a structure can be included in the group of pharmaceutical drugs, what its biological activity, chemical reactivity or toxicity will be [51–53]. Prediction of physical and/or biological properties of a given drug is usually done on the basis of a certain number of descriptors, which are appropriately tabulated (e.g., Ghose Crippen descriptors, topological pharmacophore (CATS) and others). Reverse prediction can also be made, i.e., knowing the appropriate relation between the structure of the drug with e.g., its activity, the structure of the tested drug is recognized [49,54,55].

There are only a few studies on the use of SVM and THz spectra for the identification of drugs or candidates [56,57] of drugs other than the prohibited. In [56], the THz spectra were used for recognition of traditional Chinese medicaments with an artificial neural network (ANN) and Kernel support vector machine (KSVM) with untrained linear, polynomial and Gaussian radial basis function (GRBF) kernels. KSVM outperformed ANN for all the medicaments at the constant parameters of the kernels. In our study, we applied KSVM with an adaptive kernel, which is much more suitable for practical applications. It should be also mentioned that the initial results on using KSVM for recognition of bi-heterocyclic compounds (BHCs) were presented in the conference correspondence [57] but also with constant parameters in a kernel.

In this study, we focus on the classification of 13 bi-heterocyclic compounds that have not yet entered production but are at the 'in vitro' testing stage. THz-TDS and self-optimizing SVM combined with the *k*-Nearest Neighbour (kNN) algorithm were used to identify the samples under investigation. The exact application of the algorithms is described in Section 3.

The remainder of this paper is organized as follows: Section 2 presents the experimental setup and a short description of the analyzed bi-heterocyclic compounds. The statistical tools are described in Section 3. The classification results are presented in Section 4. Finally, the conclusions are drawn in Section 5.

2. Experimental Study

A system for the recognition of pharmacological bi-heterocyclic compounds is shown in Figure 1. Its main components are a terahertz spectrometer (THz-TDS) combined with a motorized rotary sample holder (MRSH) and a computer with MATLAB software (version 2016a, 64-bit, MathWorks, Natick, MA, USA).
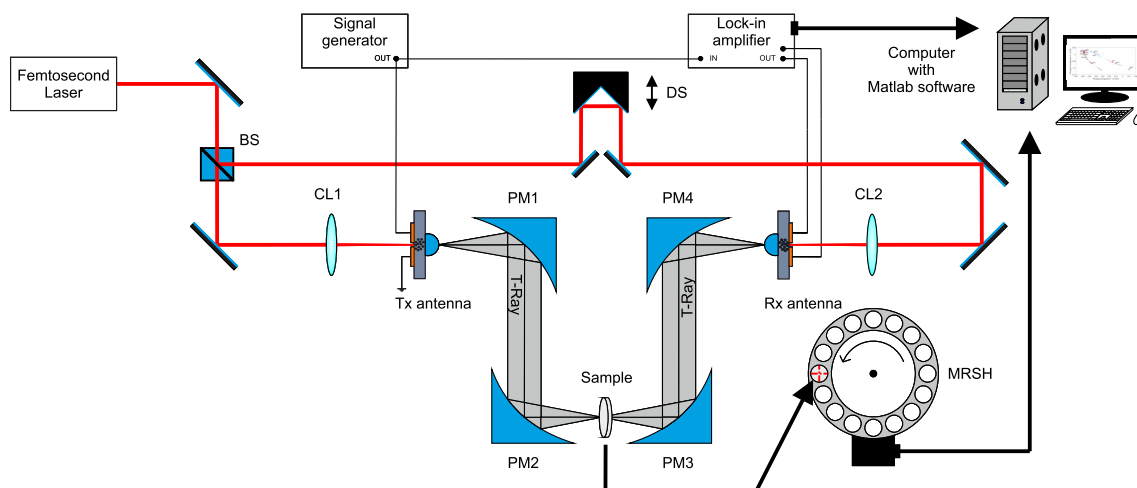
Figure 1. Terahertz system for the recognition of bi-heterocyclic compounds.

2.1. TDS-THz Spectrometer

Spectral measurements of samples were performed using the classical terahertz spectrometer THz-TDS. Femtosecond laser was characterized by an average power of 200 mW, pulse duration of 8.5 fs and pulse repetition frequency of 100 MHz. The 780 nm laser output beam is divided into two beams—excitation (EB) and sampling (PB)—by means of a light-dividing cube (BS). The first one goes to the transmitting antenna (Tx), and the second to the receiving antenna (Rx). In both cases, the same standard LT-GaAs photoconductive antennae have been used, resulting in a useful bandwidth of 2.1 THz. Both beams (EB and PB) are guided by means of mirrors and focused on the gap between antenna dipoles by means of microscopic lenses (MO). The terahertz beam path was made by means of four parabolic off-axis mirrors (PM1-4). Between PM2 and PM3, there is a rotating, numerically controlled holder for 16 test samples. The rotation is realized in such a way that, after registering the spectrum of one compound, the handle is rotated by 22.5° to the position of the next sample. Connections of signal generator, lock-in amplifier and delay line (DS) (retroprism mounted on electric linear drive) shown in Figure 1 recreate the pulse on Rx. The range of the band for the studied bi-heterocyclic compounds was limited to 10–70 cm^{-1}.

2.2. Method of Manufacture of the Samples

A single sample is a crystalline form of a bi-heterocyclic compound together with a pure form of polyethylene (PE) formed as a pill (pellet) with a diameter of 13 mm. The weight ratio of pure compound to PE is 40 mg to 360 mg, which corresponds to a percentage ratio of 10% to 90%. The reference sample is a PE pill that has been reduced to 360 mg in weight. A hydraulic press of two tonnes for two minutes is used to make pellets.

2.3. Chemical Structure of Bi-Heterocyclic Compounds

The chemical structures of all compounds tested by us, labelled as BHC1-13, are presented in Figure 2. They belong to a group of bi-heterocycles with a common core based on isothiazolpyridine derivatives, more precisely 4,6-dimethyl-2,3-dihydroisothiazolo[5,4-b]pyridin-3-on (Figure 2, last item, in blue). These are compounds inhibiting cyclooxygenases COX-1 and COX-2 [58,59] and are still in the 'in vitro' testing phase. Drugs that have similar properties but are registered and are commonly found in the pharmaceutical market are referred to as non-steroidal anti-inflammatory drugs (NSAIDs). Terahertz spectra of visible structures are presented in the further part of the study (Section 3.1, Figure 3).
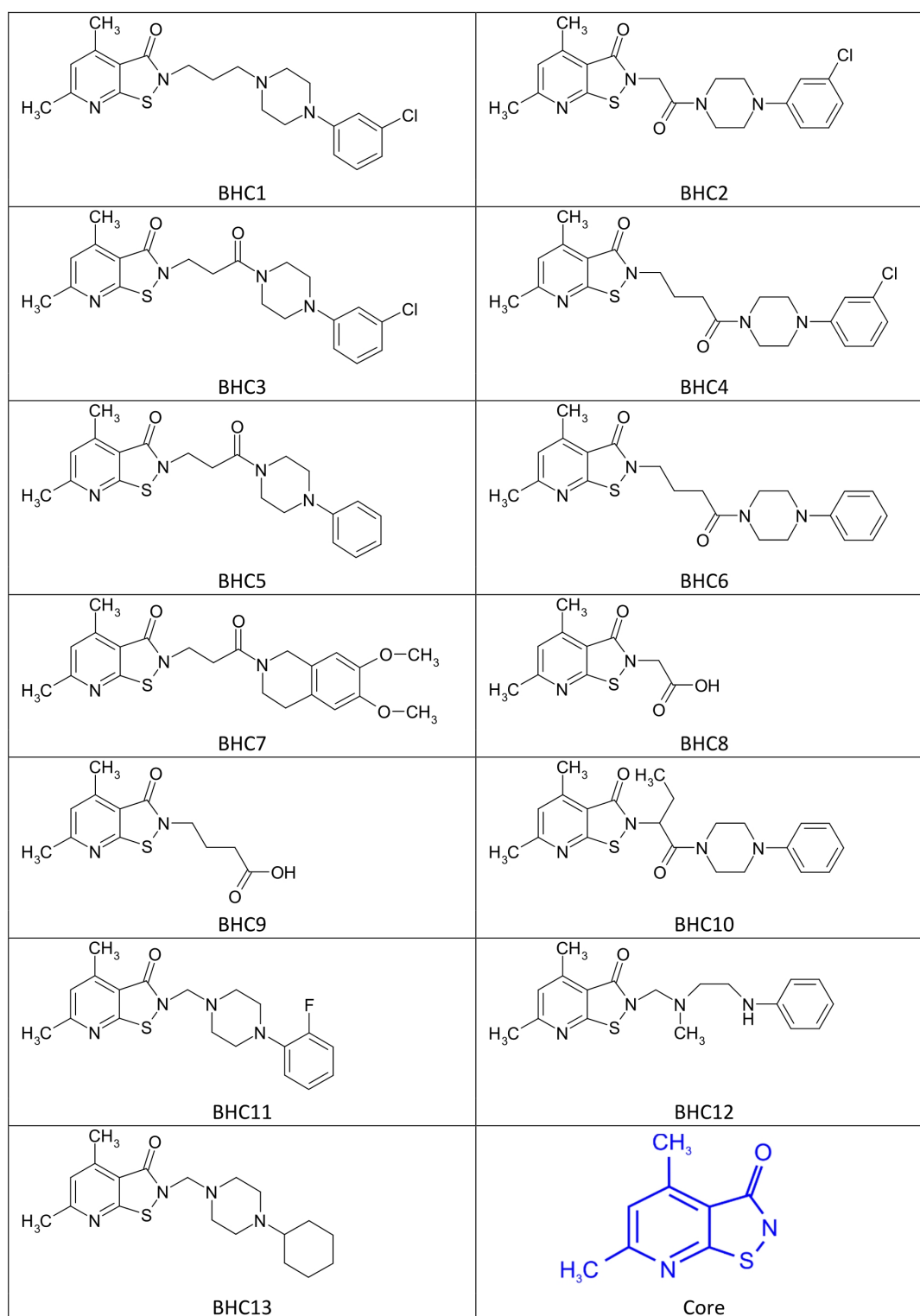
Figure 2. Chemical structures of bi-heterocyclic compounds (BHC1-13) used in the studies. The common core is marked in blue.

3. Statistical Analysis

Terahertz spectra of bi-heterocyclic compounds contain regular structures that can be analyzed with advanced machine learning methods. The aim is to design a recognition method that can uniquely recognize bi-heterocyclic compounds based on their observed THz spectra. Our approach consists

of two stages. First, relevant features are extracted from a set of observed spectra using principal component analysis (PCA) that belongs to a family of unsupervised machine learning methods. In the other stage, supervised learning methods are used to classify the features. We observed that the measured spectral samples, projected onto the space spanned by the PCA features, form regular clusters that might be separated by a nonlinear hypersurface. This motivates the usage of a nonlinear support vector machine (SVM) classifier with an adequately specified kernel function. To maximize the margin between the samples projected onto the nonlinear feature space, we propose to learn the parameters of the kernel function from training data using the cross-validation (CV) technique and the sequential quadratic programming (SQP) method [60] to minimize the misclassification error function. The concept of the proposed learning algorithm is presented below.

3.1. PCA

To design a supervised learning recognition algorithm, the observed THz spectra are partitioned into training and testing samples, e.g., according to the *n*-fold CV rule. We assume that the testing spectra can be classified into *C* disjoint classes. The examples of observed spectra are shown in Figure 3.
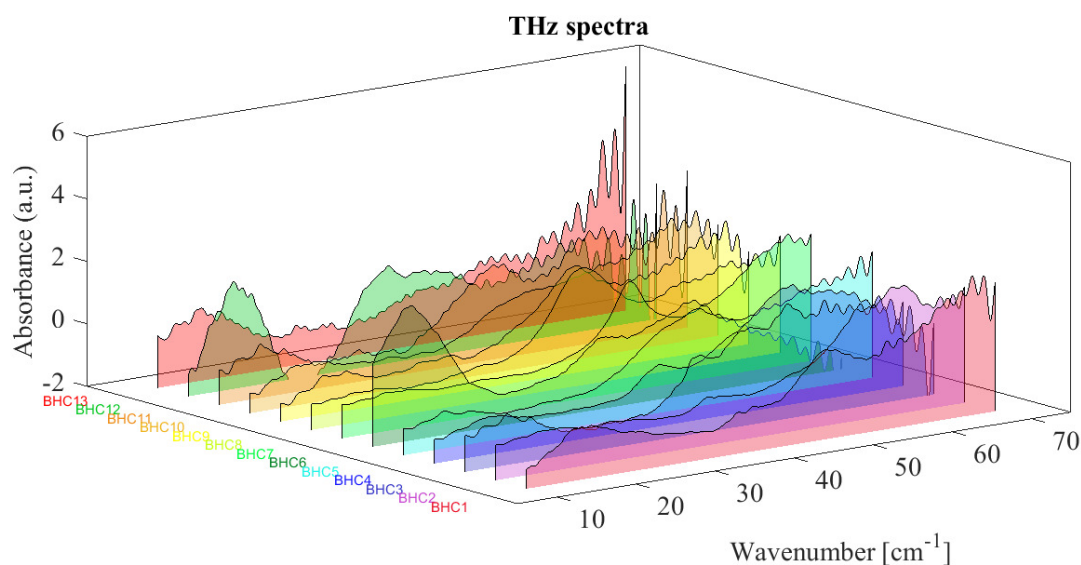

Figure 3. THz spectra of tested bi-heterocyclic compounds.

Spectral resolution of observed spectra is usually high, and each spectrum contains a lot of redundant information. Reduction of redundancy can therefore be performed with PCA. It extracts *J* orthogonal features that trace the directions in the *I*-dimensional space of observed spectra, and the samples projected onto these directions have the highest possibly variability. The features are represented by the eigenvectors associated with the *J* largest eigenvalues of the empirical covariance matrix of the training spectra. The centralized observed samples projected onto the space spanned by the orthogonal features determine the principal components (PCs). The variance of each PC is determined by the eigenvalue associated with the corresponding feature.

The scatter plots of each pair of three most significant PCs are illustrated in Figure 4 for the spectra of 13 bi-heterocyclic compounds (Figure 3). The notation of BHC1-BHC13 correspond to the chemical structures illustrated in Figure 2. The symbols "red plus" stand for the centroids of the clusters. Each class is represented by 30 samples.

(a)

(b)

(c)
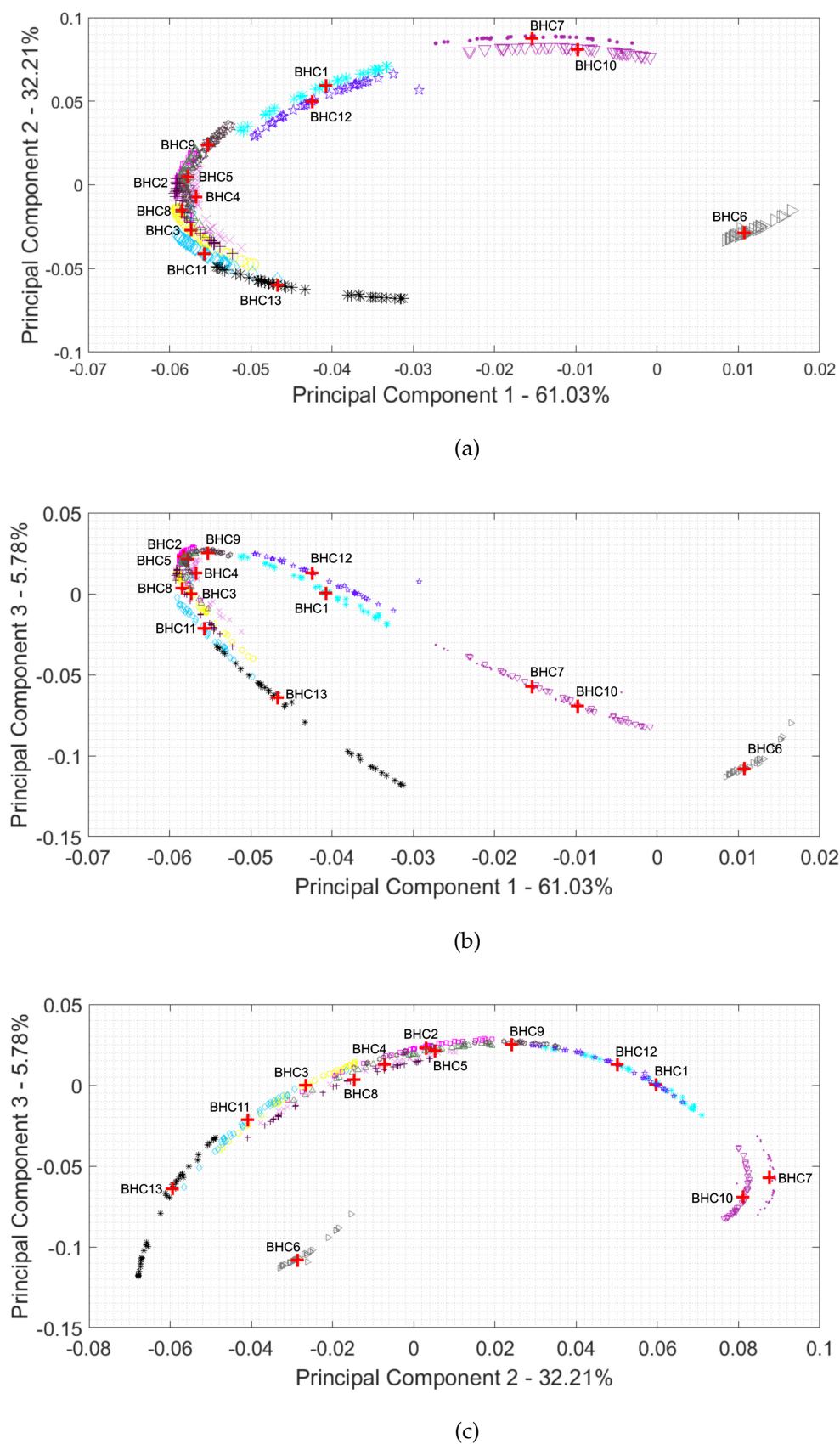
Figure 4. PCA map of thirteen bi-heterocyclic compounds: (**a**) PC1 and PC2; (**b**) PC1 and PC3; (**c**) PC2 and PC3.

PC1, PC2, and PC3 explain 61.03%, 32.21%, and 5.78% of the total variance. Thus, the explained variance by the first three PCs exceeds 99%. If $J << I$, the reduction of dimensionality is very large. In our experiment, $I = 1194$ and $J = 3$. These results confirm that the THz spectra of bi-heterocyclic compounds contain highly redundant information, and a small number of PCs is sufficient to capture more than 99% of the total variance. However, three PCs are not sufficient to perform a successful classification, despite they capture a meaningful amount of variance. In general, the variance reflects the energy of a signal, and spiky signals, despite their high magnitude, do not carry high energy, but they are informative and might be discriminant. In other words, any spikes in a spectrum do not affect the variance considerably, but they may uniquely characterize a group of bi-heterocyclic compounds. Hence, we must use a higher number of PCs than comes from the variance analysis, which will be confirmed by the experiments. A similar conclusion was given in the paper [61] for classification of spectra in laser-induced breakdown spectroscopy (LIBS), and this behavior was also experimentally justified in the work [62].

3.2. KSVM

The samples shown in Figure 4 form consistent but non-Gaussian clusters that have arc structures, and some clusters cannot be separated with linear hyperplanes. This gives us empirical evidence that nonlinear hypersurfaces should be used to separate the groups of samples. To tackle this problem computationally, the observed samples can be projected onto the high-dimensional feature space using a nonlinear mapping $\Phi : \mathbb{R}^J \to \mathbb{R}^D$, where $D >> J$. The feature space can also be infinite dimensional Hilbert space of real-value functions with inner product $< \cdot, \cdot >$ and norm $|| \cdot ||$. The aim is to find a nonlinear mapping Φ such that the samples $\{\Phi(z_r)\}$ are linearly separably in the feature space.

This is a basic concept of the nonlinear or Kernel Support Vector Machine (KSVM) that is successfully used in many areas of science [63,64]. The fundamental version of this classifier performs binary classification by finding the hyperplane $\mathcal{H} = \{x : w^T\Phi(x) + b = 0\}$ that best separates both classes in the feature space. The vector $w \in \mathbb{R}^D$ is normal to $\mathcal{H}$, and $\frac{b}{||w||_2}$ is the perpendicular distance from $\mathcal{H}$ to the origin in $\mathbb{R}^P$. The data points located closest to $\mathcal{H}$ are referred to as the support vectors (SVs). The best separating hyperplane should be selected in such a way that the distance between it and the point determined by any support vector in both classes is maximized.

Let $\{x_r, y_r\}$ for $r = 1, \ldots, R$ be a training dataset, where $y_r \in \{-1, 1\}$ is an indicator of the class to which x_r belongs. The task of finding the best separating hyperplane boils down to the constrained quadratic programming (QP) problem:

$$\min_{\lambda} \left(\frac{1}{2}\lambda^T H\lambda - e^T\lambda \right), \quad \text{s.t.} \quad y^T\lambda = 0 \quad \text{and} \quad \forall r : C_\xi \geq \lambda_r \geq 0, \tag{1}$$

where $H = [h_{rs}] \in \mathbb{R}^{R \times R}$ with $h_{rs} = y_r y_s \Phi^T(x_r)\Phi(x_s)$, $y = [y_r] \in \mathbb{R}^R$, $e = [1, \ldots, 1] \in \mathbb{R}^R$, and $C_\xi \geq 0$ is the soft-margin penalty parameter that is introduced to tackle the existence of inevitable outliers. The solution λ_* to the optimization problem in formula (1) uniquely determines the parameters w_* and b_* in the separating hyperplane $\mathcal{H}_* = \{x : w_*^T\Phi(x) + b_*\}$ that implies the decision rule for testing samples in J-dimensional space.

The mapping $\Phi(x_r)$ transforms the samples x_r from a J-dimensional to a very high-dimensional space. Fortunately, the inner product $\Phi^T(x_r)\Phi(x_s)$ does not have to be computed explicitly. Due to the Mercer's theorem, the inner product $\Phi^T(x_r)\Phi(x_s)$ can be computed by the kernel function:

$$k_{rs} = k(x_r, x_s) = \Phi^T(x_r)\Phi(x_s), \tag{2}$$

where the matrix $K = [k_{rs}] \in \mathbb{R}^{R \times R}$ is symmetric and semi positive-definite. Thus, the QP problem in expressed by (1) is convex and can be easily solved by many solvers using, e.g., the active set or interior point algorithm. In the experiments, we used the Sequential Minimal Optimization (SMO) algorithm.

The kernel function in (2) can take various forms. For example:

- d-**degree polynomial:** $k_1(z_r, z_s) = (z_r^T z_s + \beta_0)^d$, where $d \geq 1$ and $\beta_0 \geq 0$,
- **Gaussian Radial Basis Functions (GRBF):** $k_2(z_r, z_s) = \exp\{\frac{||z_r - z_s||^2}{2\sigma^2}\}$, where $\sigma^2 > 0$,
- **two-layer perceptron:** $k_3(z_r, z_s) = \tanh(\beta_0 z_r^T z_s + \beta_1)$, where $\beta_0, \beta_1 > 0$.

Note that, if $d = 1$ and $\beta_0 = 0$ in $k_1(z_r, z_s)$, then KSVM with this kernel simplifies to the standard SVM. When $d = 2$ and $J = 2$, i.e., $z_r = [z_{1r}, z_{2r}]^T$ and $z_s = [z_{1s}, z_{2s}]^T$, we have $\Phi(z_r) = [z_{1r}^2, z_{2r}^2, \sqrt{2}z_{1r}z_{2r}, \sqrt{\beta_0}z_{1r}z_{2r}, \sqrt{2\beta_0}z_{1r}, \sqrt{2\beta_0}z_{2r}, \beta_0]^T \in \mathbb{R}^7$. For a larger value of J, the dimensionality of the feature space is much bigger even for a second-degree polynomial. Fortunately, due to the kernel assumption, the mapping Φ is never computed explicitly in the whole KSVM algorithm.

The curvature of the nonlinear separating hypersurface can be controlled by the kernel function and its related parameters. It is obvious that the curvature should be adapted to data, which implies that the kernel parameters are learnt from training data. Let $p = \{d, \beta_0, \beta_1, \sigma^2, C_\xi\}$ be a vector of the parameters to be estimated. To estimate the vector p, the n-fold CV was applied separately to each class.

The main criterion used for testing and validating the model is the Misclassification Ratio (MCR) that is defined as $\Psi = \{\sharp t : y_t - y_t^{(true)} \neq 0\}$, where y_t is a binary variable expressed by the decision rule in KSVM, and $y_t^{(true)}$ is the true label of the class to which the t-th testing sample belongs to. Note that the function Ψ is parameter-dependent. Thus, the problem of data-driven estimation of the parameters can be formulated by the following constrained optimization problem:

$$\min_z \Psi(z), \quad \text{s.t.} \quad \forall l : c_l(p) \geq 0, \tag{3}$$

where $c_l(z)$ are constraints.

The function $\Psi(z)$ is not differentiable with respect to z. To solve problem (3), derivative-free methods must be used, such as evolutionary strategies, metaheuristics, or direct search methods [65–67]. In our experiments, we chose the Nelder–Mead (NM) method [68] that belongs to a family of direct search strategies for multi-variate unconstrained optimization. It generates a sequence of simplexes for which the objective function at their vertices decreases its value. Starting from $(J + 1)$-dimensional simplex in the space $\mathbb{R}^J$, the vertices of the current simplex are updated according to an expansion–shrink strategy that searches for reflection–contraction points that are used in decision rules to expand or shrink the current simplex. The NM method is basically used for unconstrained nonlinear optimization, but it can be applied to problem (3) because the constraints in all of the above-mentioned kernel functions are strictly positive and can be expressed by the following exponential mapping: $z = \exp\{\tilde{z}\}$, where $\tilde{z}$ is an unconstrained vector of search variables in the NM method.

The standard KSVM is a binary classifier. Thus, it can be directly applied to binary decision problems, e.g., to decide whether a given THz spectrum represents a given bi-heterocyclic compound or not. For multiple bi-heterocyclic compounds, one-vs.-all classification strategy can be applied. Hence, we use as many KSVM classifiers as there are classes, and each classifier is trained to recognize one bi-heterocyclic compound against the rest. Then, in the testing process, a testing sample is verified separately by each classifier. Note that this methodology incurs a high computational cost, but it offers many additional advantages. For example, if a testing sample cannot be identified by any trained classifier, we can apply another classifier only to the unrecognized sample, or we can assign this sample to some unknown class. The samples that fall to such a new class are candidates for representing a new bi-heterocyclic compound, and they are further classified with the k-NN classifier. Similarly, if a testing sample is recognized by more than one classifier, we can also repeat the classification with another, more efficient, classifier. This approach is particularly useful in practice, when outliers or other perturbations occur.

4. Classification Results

Various experimental tests are performed to demonstrate efficiency of the proposed approach in bi-heterocyclic compounds (BHC) recognition. In the first stage, the KSVM with an additional class and the non-adaptive GRBF kernel was used, where $\sigma^2 = C_\xi = 1$. In the second stage, the parameter σ^2 was learnt automatically from the observed spectra. In the preliminary tests, we observed that the soft-margin C_ξ has a negligible effect on the accuracy of recognition, hence we assumed $C_\xi = 1$ in the further tests. Neglecting the updates for the variables w and b in the primal QP problem, problem (3) boils down to one variable nonlinear optimization problem.

In the experiments, we used 390 absorbance spectra registered by the THz-TDS spectrometer for 13 bi-heterocyclic compounds that are described in Section 2.3. The spectral resolution amounts to 1194 subbands, and each compound is represented by 30 spectra. All of the originally registered spectra were preprocessed with PCA where the number of PCs was selected in the range (1, 40). The training was performed on 12 selected compounds labeled according to Figure 2. The 13th compound is regarded as unknown, and its spectra will be referred to as outer-class samples which will be used only for testing. Hence, they form an additional (extra) class (labeled as 13). Moreover, the testing samples are not taken from the same spectra dataset that is used for training. Due to the long-term drift effects and other slowly-varying conditions of measurements, the spectra used for testing are measured several days later than the other samples. The testing set consists of 30 samples of mixed classes, including the outer-class. We analyze 11 scenarios with an unknown compound, changing the number of spectra from the outer-class in the testing dataset, from zero to 100% with the step of 10%. To check the repeatability of the results, each scenario was repeated 13 times—each time one compound was neglected. It means that the training process was performed for each scenario with one (each time different) missing class, and, in the testing, the samples from the missing class should be classified to the extra class.

Since the objective function in (3) is non-convex with respect to z, the NM method may fall in local minima. To minimize such a risk, multi-start initialization is necessary. Thus, each experiment with data-driven parameter estimation was also repeated with respect to the multi-start NM initialization. Finally, the MCR is averaged separately for each scenario.

The proposed algorithm was coded in MATLAB R2016a and run on the machine equipped with two CPU Intel Xeon(R) E5-2699V3 (Santa Clara, CA, USA), 2.30 GHz.

4.1. Untrained GRBF Kernel

The averaged accuracy of BHC recognition obtained by the KSVM with the untrained GRBF (fixed $\sigma^2 = 1$) versus the number of PCs is shown in Figure 5. This algorithm yields the mean accuracy of recognition greater or equal to 90% only for the PCs in the range [10, 21]. The accuracy exceeds 97% for three values of PCs ($J = 13, 18, 19$) with the highest rate equal to $\eta_{avg} = 97.93\%$ for $J = 18$. For other numbers of PCs in the range [1, 40], the accuracy is below 90%.

Figure 6 presents the mean accuracy versus PCs in the range [13, 20] (non-blue bars in Figure 5) at various percentage levels of contribution of the outer-class bi-heterocyclic compound (from zero to 100% at the step of 10%). The accuracy changes considerably with the number of PCs and the number of outer-class samples. The accuracy reaches 100% for $J = 13$, when all the test samples are known, i.e., they belong to the classes that are used for training. If outer-class samples occur and a low value of PCs is used, the efficiency of recognition falls down noticeably. The accuracy averaged over the contribution of outer-class samples is about 95.13%. For $J \in [20, 40]$, we observed a reverse behaviour. When all the testing samples belong to an extra class, $\eta_{avg} = 100\%$, while, if some testing samples from other classes occur, the accuracy drops considerably. For $J = 20$ and all the known samples, $\eta_{avg} = 93.59\%$, while we have only $\eta_{avg} = 18.46\%$ for $J = 40$. For $J = 13$ and $J = 18$, the accuracy was similar, i.e., 97.88% and 97.93%, respectively. Thus, the choice of the parameter J may be conditioned by our expectation—better recognition of unknown compounds or better classification of testing samples that are closely related to training samples. The optimal choice seems to point at

$J = 18$ with the worst accuracy of 96.92%, and the best—99.23% if all the testing samples represent a new compound. When $J = 18$, the explained variance amounts to 99.9984%, but the MCR at the level of 3% (in the worst case) and its variability with the number of PCs in quite a narrow range is not fully satisfactory in applications to pharmaceutical industry and medal chemistry.
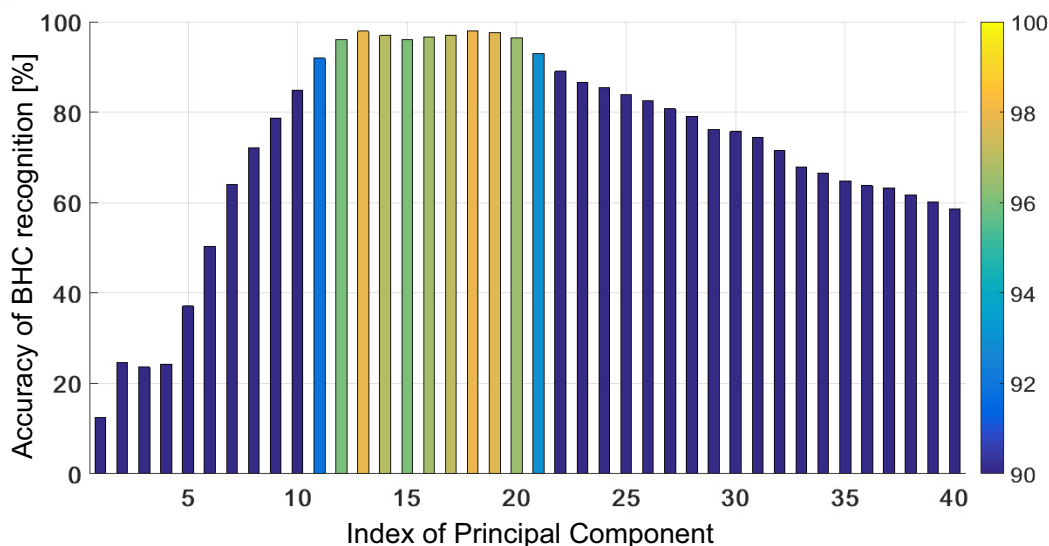
Figure 5. Accuracy of BHC recognition (average) versus PCs for the KSVM with the untrained GRBF.

The mean training time for the discussed case ($J = 18$) is 391.8 ms with the standard deviation (STD) of 3 μs. The testing time obviously depends linearly on the number of testing samples, and can be determined by the regression line $ET(n_t) = 3.9n_t + 1.4$ [ms] with the confidence interval of 95%, where n_t is the number of testing samples. The overall time for testing 30 samples takes about 120 ± 1.3 ms.
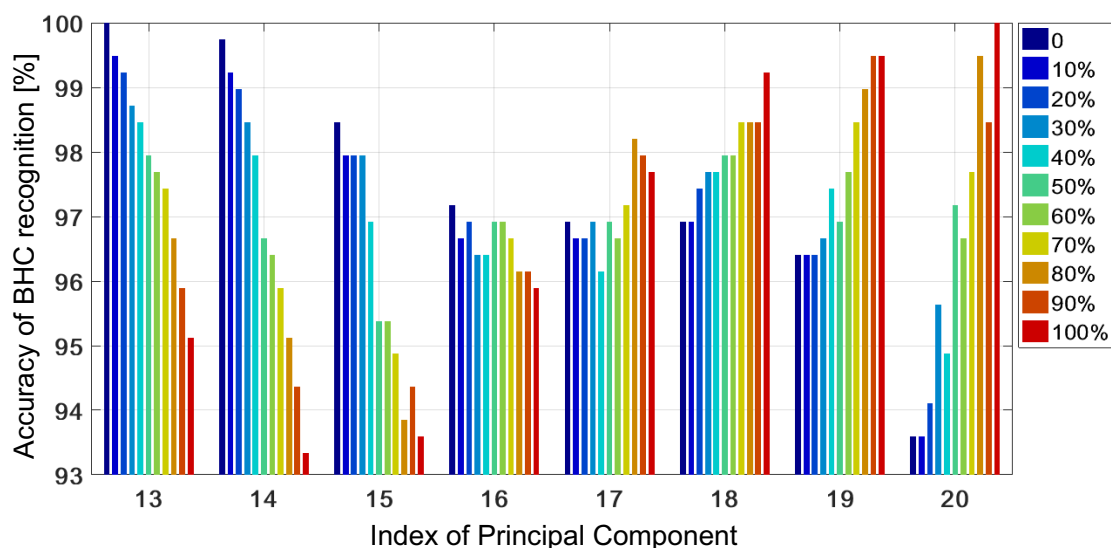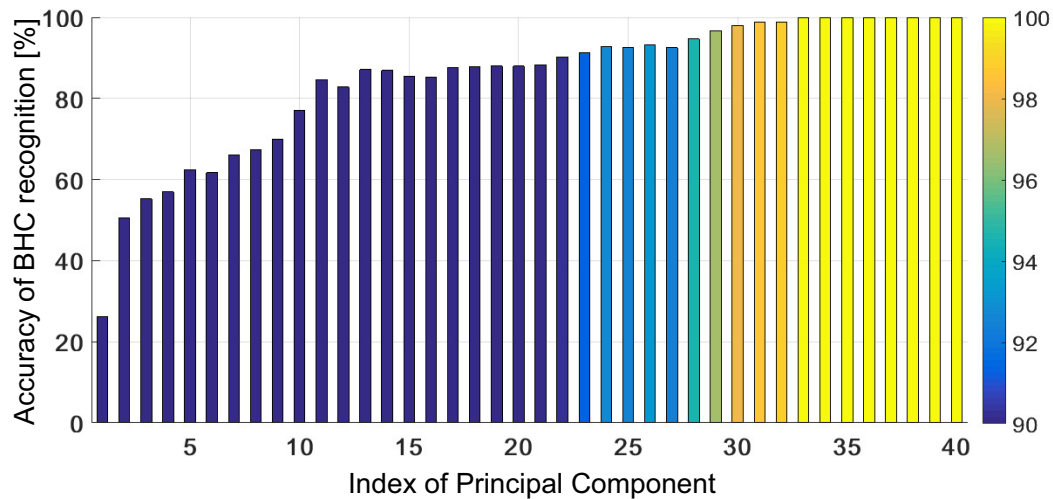

Figure 6. Accuracy of BHC recognition obtained with non-adaptive KSVM versus the number of PCs (from 13 to 20) at various contributions of outer-class samples in the test set.

4.2. Trained GRBF Kernel

The results obtained with the adaptive KSVM are shown in Figure 7 versus the number of PCs. For this approach, we were able to obtain 100% mean accuracy of BHC recognition for $J \geq 34$. The

tests are performed for the 2-fold CV for training the kernel parameters and 30 initialization points in the NM algorithm. The proposed algorithm is resistant to the amount of outer-class samples in the testing set. Reducing the dimensionality from 1194 subbands to 34 PCs, 99,9957% of the total variance is preserved.

Figure 7. Accuracy of BHC recognition (average) obtained with adaptive KSVM versus the number of PCs. The test was performed using 2-fold CV and 30 initialization points for the NM algorithm.

For $J \in [1, 21]$, the mean accuracy does not exceed 90 % but when $J \geq 30$, the advantage of the KSVM with the trained GRBF is considerable with respect to the untrained case. When the testing samples contain a new BHC, the KSVM with the trained GRBF can adapt to each analyzed case when $J \geq 25$, and is able to correctly classify all the testing samples, if they do not contain outer-class samples—see the results presented in Figure 8. Moreover, in this respect, it outperforms the KSVM with the untrained GRBF considerably—compare the results presented in Figures 6 and 8.
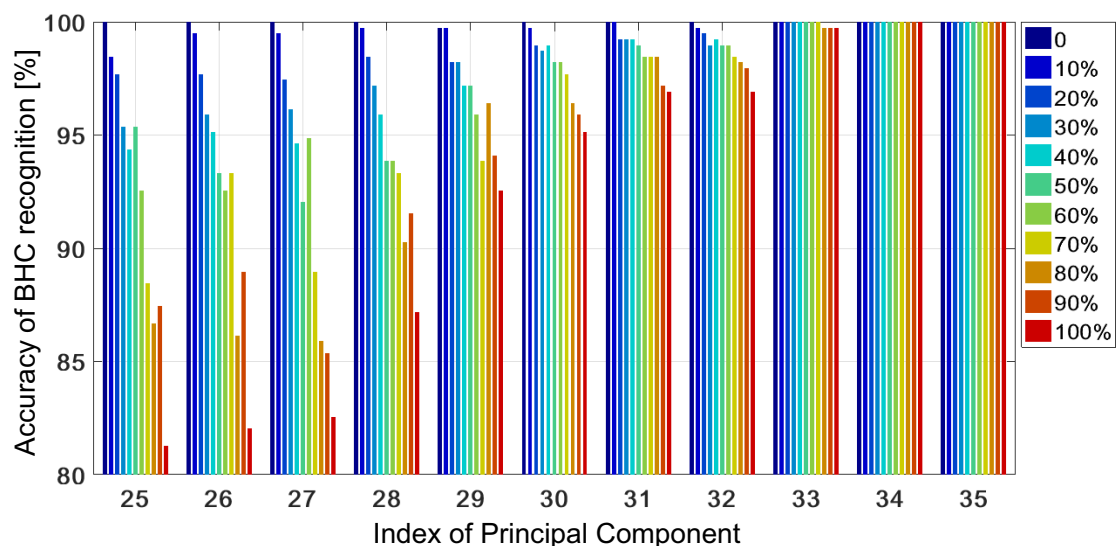
Figure 8. Accuracy of BHC recognition obtained with adaptive KSVM versus the number of PCs (from 25 to 35) at various contributions of outer-class samples in the test set.

In the range $J \in [30, 40]$, the accuracy is nearly constant with CV folds used for training the kernel. However, the number of CV folds has a great impact on the training time, which is illustrated

in Figure 9. The results are obtained for $J = 35$ and 30 initialization points in the NM algorithm, where 99.9959% of the total variance is explained. The median of training time of KSVM combined with running PCA changes from 6.6822 s. (2-fold CV) to 26.2272 s. (10-fold CV).
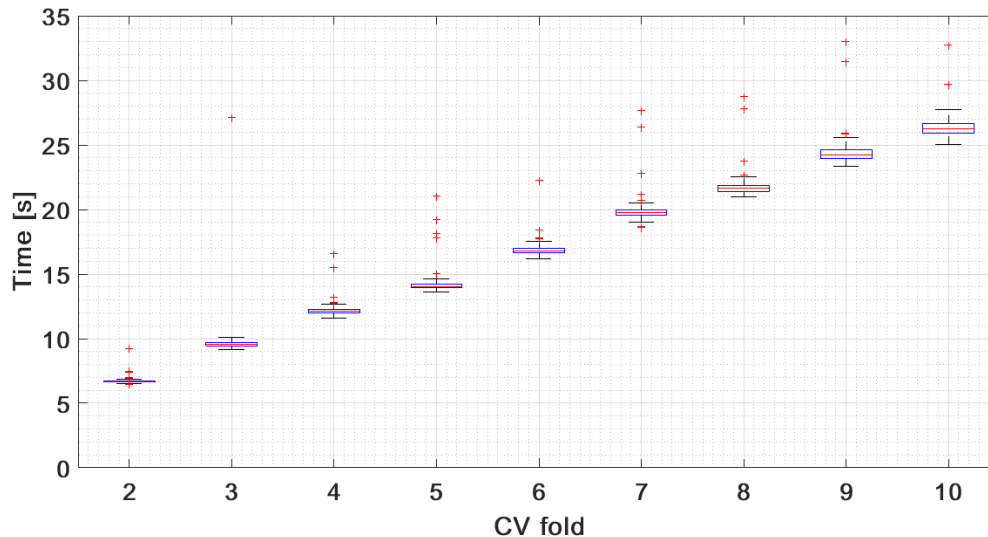
Figure 9. Time-training statistics (box plots) obtained for the adaptive KSVM versus the number of CV folds. The test was performed at 35 PCs and 30 initialization points in the NM algorithm.

We also analyzed how an increase in the number of CV folds affects a minimal number of PCs needed to reach the accuracy of 100%. For the CV with 5–10 folds, 100% accuracy is possible at $J = 33$. The time profit obtained from a decrease J by two at 5-fold CV with respect to $J = 35$ at the 2-fold CV does not compensate the total time lost of 7.1333 s. Another parameter that considerably affects the accuracy and runtime is the number of initialization points in the NM algorithm. Figure 10 presents the accuracy vs. the parameter J at various numbers of initialization points [5, 10, 15, 20, 30, 50].
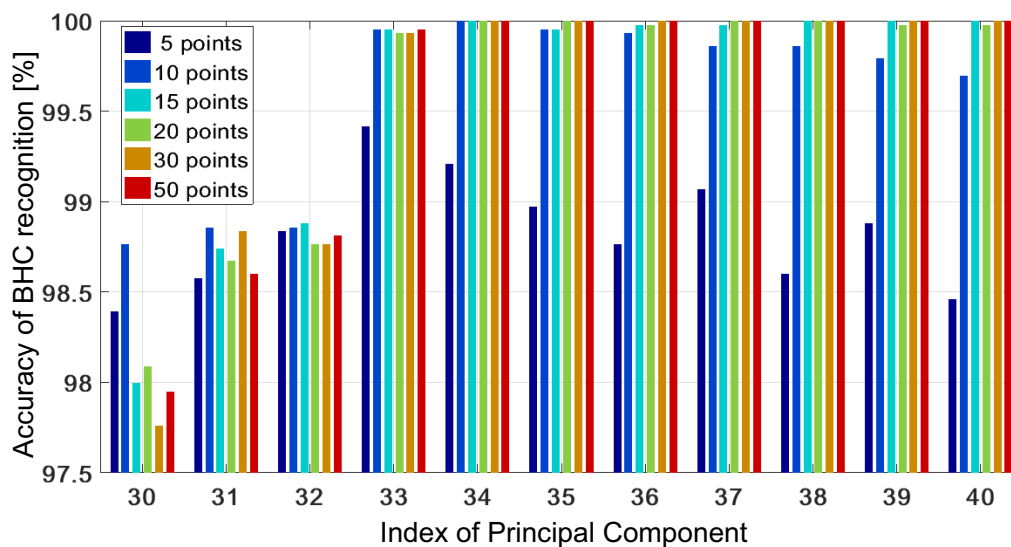
Figure 10. The accuracy of BHC recognition versus the number of PCs for various numbers of initialization points in the NM algorithm.

If the number of these points is too small, there is a higher probability that the solution to the problem (3) is suboptimal. Figure 10 shows that the accuracy of 100% is not obtained in the whole

range of PCs ($J \in [30, \ 40]$) when only five points are used. However, when 10 points are used, all of the test samples are correctly recognized at $J = 34$. When the number of the points exceeds 30, the highest accuracy is observed in the desired range of PCs.

The training time of the KSVM with an adaptive GRBF versus the number of initialization points at $J = 35$ and 2-folds CV is listed in Table 1.

Table 1. The mean training time [s] of the KSVM with an adaptive GRBF versus the number of initialization points in the NM algorithm.

No. initialization points:	5	10	15	20	30	50
Mean training time [s]:	4.01	4.84	5.34	5.97	6.7	8.22

The training time increases exponentially with the number of initialization points in the NM algorithm, and it can be modelled by the regression curve (with a confidence level of 95%): $ET_{(train)}(s) = -2.371 \exp\{-83.45 \times 10^{-3}s\} + 5.349 \exp\{8.371 \times 10^{-3}s\}$, [s], where s is the number of the points. Considering the obtained accuracy, training time, and non-convexity of the NM algorithm, the optimal number of the initialization points seems to be 30.

For the KSVM with the trained GRBF with $J = 35$, the mean training time amounts to 6.71 ± 0.18 s, and its median is 6.68 s. Similarly as for the untrained kernel, the testing time can be approximated by a linear function $ET(n_t) = 0.369n_t + 0.536$ [ms] with the confidence interval of 95%, where n_t is the number of testing samples. The overall time for testing 30 samples takes about 11.6 ms.

5. Conclusions

The method for classification of bi-heterocyclic compounds developed by us can be successfully used in medical chemistry and pharmaceutical industry. The compound spectra (BHC1-13) recorded with the THz-TDS terahertz spectrometer provide enough information, so that the compounds can be classified unambiguously by means of machine learning methods. It should be remembered that the compounds selected for the study have the same core (isothiazolpyridine derivatives) in the chemical structure. Assigning the spectrum of an unknown sample to a specific group can be performed with many classification methods. However, the KSVM trained separately to each class gives us a possibility to assign a testing sample to an unknown group which can be analyzed with other methods, e.g., with the *k*-NN classifier.

In medical chemistry, the presence of new compounds is very desirable and it is important to determine quickly and unequivocally whether it is a new structure or a known, previously synthesised compound. The KSVM with the untrained kernel does not meet our expectations because, depending on the percentage of unknown content of samples in the test set, for the best case (for 18 PCs), allows for one to four misclassifications per every 100 samples.

This problem was solved by using the KSVM with the trained GRBF. The proposed algorithm assumes the splitting of the classification problem of C known compounds into C binary problems, which allows the creation of an additional (extra) class $C + 1$ to which a new, previously unknown sample can be assigned. In this way, using e.g., cross-validation and and the derivative-free method for training the kernel, the algorithm adjusts the parameters on its own (*w*, *b* and the kernel scaling function σ) optimizing the hyperplane separating individual classes. After creating a single decision rule using samples from cross-validation, the test samples of any bi-heterocyclic compounds are classified. When carrying out the testing C times, three results are obtained. The sample can be assigned an index of a known class, an index of an unknown class $(C + 1)$ or several indexes of known classes. In the third case, the final classification decision is taken with the help of *k*-NN with the Euclidean metric. For the studied group of BHC1-13 compounds, we achieved the efficiency of classification $\eta = 100\%$ regardless of the number of unknown bi-heterocyclic compound samples in the test set. The optimal solution is provided by the use of PCA pre-processing with 35 components, 2-fold cross-validation and 30 starting points in the NM algorithm. The training time of our algorithm

is much longer than for the KSVM with the untrained Gaussian kernel (6.7 s vs. 390 ms), but may be partially compensated by the time of sample recognition, especially when loading long series of spectra. In the range of 1–30 test samples, we gain the time spent on sample classification from about 6 (comparison for 1 sample in the test set: 4.675 ms vs. 0.792 ms) to more than 10 times (comparison for 30 samples: 120 ms vs. 11.6 ms).

In summary, the results show that the proposed THz-TDS system can perfectly classify all the BHCs, regardless of the number of outer-class samples in the testing set. In this study, we focused on classification of the BHCs having the common core and different molecular tails (linkers with functional groups) that inhibit cyclooxygenases COX-1 and COX-2. Taking into account two opposite measurement cases with the testing samples—all belonging to the space of training samples or all are outer-class samples—we can conclude that, in the former case, it is possible to recognize all molecular tails linked to the same core. In the other case, we observed that the proposed sensor system can correctly recognize new tails and classify them to a new group, even if their structures are similar to those used for training (e.g., BHC1 and BHC2 in Figure 1). It is therefore the method sensitive to any changes introduced to the BHC. The analysis of BHCs with other cores will be performed in our future research.

References

1. Eicher, T.; Hauptmann, S.; Speicher, A. *The Chemistry of Heterocycles: Structures, Reactions, Synthesis, and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
2. Joule, J.A.; Mills, K. *Heterocyclic Chemistry at a Glance*; John Wiley & Sons: West Sussex, UK, 2012.
3. Yanagimoto, K.; Lee, K.G.; Ochi, H.; Shibamoto, T. Antioxidative activity of heterocyclic compounds found in coffee volatiles produced by Maillard reaction. *J. Agric. Food Chem.* **2002**, *50*, 5480–5484. [CrossRef] [PubMed]
4. Agranat, I. Pharmaceutical Substances: Syntheses, Patents and Applications of the Most Relevant AIPs. *Synthesis* **2009**, *2009*, 2116. [CrossRef]
5. Lamberth, C.; Dinges, J. *Bioactive Heterocyclic Compound Classes: Pharmaceuticals*; Wiley-VCH: Weinheim, Germany, 2012.
6. Kujawski, J.; Janusz, A.; Kuźma, W.; Ożarowski, M.; Kujawski, R. Biologicznie ważne substancje zawierające pierścień pirolu–działanie i występowanie. *Farm Pol.* **2010**, *66*, 117–125.
7. Selvam, T.P.; James, C.R.; Dniandev, P.V.; Valzita, S.K. A mini review of pyrimidine and fused pyrimidine marketed drugs. *Res. Pharm.* **2012**, *2*, 4.
8. Selvam, T.P.; Kumar, P.V. Quinazoline marketed drugs. *Res. Pharm.* **2011**, *1*, 1–21.
9. Majumdar, K.C.; Chattopadhyay, S.K. *Heterocycles in Natural Product Synthesis*; Wiley-VCH: Weinheim, Germany, 2011.
10. Alvarez-Builla, J.; Vaquero, J.J.; Barluenga, J. *Modern Heterocyclic Chemistry, 4 Volume Set*; Wiley-VCH: Weinheim, Germany, 2011; Volume 2.
11. Paudel, A.; Raijada, D.; Rantanen, J. Raman spectroscopy in pharmaceutical product design. *Adv. Drug Deliv. Rev.* **2015**, *89*, 3–20. [CrossRef] [PubMed]

12. Ricci, C.; Nyadong, L.; Yang, F.; Fernandez, F.M.; Brown, C.D.; Newton, P.N.; Kazarian, S.G. Assessment of hand-held Raman instrumentation for in situ screening for potentially counterfeit artesunate antimalarial tablets by FT-Raman spectroscopy and direct ionization mass spectrometry. *Anal. Chim. Acta* **2008**, *623*, 178–186. [CrossRef]

13. Jamrógiewicz, M. Application of the near-infrared spectroscopy in the pharmaceutical technology. *J. Pharm. Biomed. Anal.* **2012**, *66*, 1–10. [CrossRef]

14. Igne, B.; Ciurczak, E.W. *Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy*; CRC Press: Boca Raton, FL, USA, 2014.

15. De Beer, T.; Burggraeve, A.; Fonteyne, M.; Saerens, L.; Remon, J.P.; Vervaet, C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int. J. Pharm.* **2011**, *417*, 32–47. [CrossRef]

16. Bressolle, F.; Audran, M.; Pham, T.N.; Vallon, J.J. Cyclodextrins and enantiomeric separations of drugs by liquid chromatography and capillary electrophoresis: Basic principles and new developments. *J. Chromatogr. Biomed. Sci. Appl.* **1996**, *687*, 303–336. [CrossRef]

17. Phadnis, N.V.; Cavatur, R.K.; Suryanarayanan, R. Identification of drugs in pharmaceutical dosage forms by X-ray powder diffractometry. *J. Pharm. Biomed. Anal.* **1997**, *15*, 929–943. [CrossRef]

18. Rösner, P.; Junge, T.; Westphal, F.; Fritschi, G. *Mass Spectra of Designer Drugs 2015*; Wiley-VCH: Weinheim, Germany, 2015.

19. Hummel, D.; Löffler, D.; Fink, G.; Ternes, T.A. Simultaneous determination of psychoactive drugs and their metabolites in aqueous matrices by liquid chromatography mass spectrometry. *Environ. Sci. Technol.* **2006**, *40*, 7321–7328. [CrossRef] [PubMed]

20. Barras, J.; Althoefer, K.; Rowe, M.; Poplett, I.; Smith, J. The emerging field of medicines authentication by nuclear quadrupole resonance spectroscopy. *Appl. Magn. Reson.* **2012**, *43*, 511–529. [CrossRef]

21. Barras, J.; Murnane, D.; Althoefer, K.; Assi, S.; Rowe, M.D.; Poplett, I.J.; Kyriakidou, G.; Smith, J.A. Nitrogen-14 nuclear quadrupole resonance spectroscopy: A promising analytical methodology for medicines authentication and counterfeit antimalarial analysis. *Anal. Chem.* **2013**, *85*, 2746–2753. [CrossRef] [PubMed]

22. Buffler, A. Contraband detection with fast neutrons. *Radiat. Phys. Chem.* **2004**, *71*, 853–861. [CrossRef]

23. Buffler, A.; Tickner, J. Detecting contraband using neutrons: Challenges and future directions. *Radiat. Meas.* **2010**, *45*, 1186–1192. [CrossRef]

24. AlNabooda, M.O.; Shubair, R.M.; Rishani, N.R.; Aldabbagh, G. Terahertz spectroscopy and imaging for the detection and identification of illicit drugs. In Proceedings of the Sensors Networks Smart And Emerging Technologies (Senset), Beirut, Lebanon, 12–14 September 2017; pp. 1–4.

25. Puc, U.; Abina, A.; Rutar, M.; Zidanšek, A.; Jeglič, A.; Valušis, G. Terahertz spectroscopic identification of explosive and drug simulants concealed by various hiding techniques. *Appl. Opt.* **2015**, *54*, 4495–4502. [CrossRef] [PubMed]

26. Federici, J.F.; Schulkin, B.; Huang, F.; Gary, D.; Barat, R.; Oliveira, F.; Zimdars, D. THz imaging and sensing for security applications—Explosives, weapons and drugs. *Semicond. Sci. Technol.* **2005**, *20*, S266–S280. [CrossRef]

27. Burnett, A.D.; Cunningham, J.E.; Davies, A.G.; Dean, P.; Linfield, E.H. Terahertz frequency spectroscopy and its potential for security applications. In *Infrared and Raman Spectroscopy in Forensic Science*; Wiley Online Library: West Sussex, UK, 2012; pp. 295–314.

28. Kawase, K.; Ogawa, Y.; Watanabe, Y.; Inoue, H. Non-destructive terahertz imaging of illicit drugs using spectral fingerprints. *Opt. Express* **2003**, *11*, 2549–2554. [CrossRef]

29. Kato, M.; Tripathi, S.R.; Murate, K.; Imayama, K.; Kawase, K. Non-destructive drug inspection in covering materials using a terahertz spectral imaging system with injection-seeded terahertz parametric generation and detection. *Opt. Express* **2016**, *24*, 6425–6432. [CrossRef]

30. Shen, Y.C.; Taday, P.F. Development and application of terahertz pulsed imaging for nondestructive inspection of pharmaceutical tablet. *IEEE J. Sel. Top. Quantum Electron.* **2008**, *14*, 407–415. [CrossRef]

31. Ho, L.; Cuppok, Y.; Muschert, S.; Gordon, K.C.; Pepper, M.; Shen, Y.; Siepmann, F.; Siepmann, J.; Taday, P.F.; Rades, T. Effects of film coating thickness and drug layer uniformity on in vitro drug release from sustained-release coated pellets: A case study using terahertz pulsed imaging. *Int. J. Pharm.* **2009**, *382*, 151–159. [CrossRef] [PubMed]

32. Felton, L.A.; Porter, S.C. An update on pharmaceutical film coating for drug delivery. *Expert Opin. Drug Deliv.* **2013**, *10*, 421–435. [CrossRef] [PubMed]

33. May, R.K.; Evans, M.J.; Zhong, S.; Warr, I.; Gladden, L.F.; Shen, Y.; Zeitler, J.A. Terahertz in-line sensor for direct coating thickness measurement of individual tablets during film coating in real-time. *J. Pharm. Sci.* **2011**, *100*, 1535–1544. [CrossRef] [PubMed]

34. Zeitler, J.A.; Shen, Y.C. Industrial applications of terahertz imaging. In *Terahertz Spectroscopy and Imaging*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 451–489.

35. Taday, P.F.; Bradley, I.; Arnone, D.; Pepper, M. Using terahertz pulse spectroscopy to study the crystalline structure of a drug: A case study of the polymorphs of ranitidine hydrochloride. *J. Pharm. Sci.* **2003**, *92*, 831–838. [CrossRef]

36. Taday, P.F. Applications of terahertz spectroscopy to pharmaceutical sciences. *Philos. Trans. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.* **2003**, *362*, 351–364. [CrossRef]

37. Kojima, S.; Shibata, T.; Igawa, H.; Mori, T. Broadband terahertz time-domain spectroscopy: Crystalline and glassy drug materials. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Tsukuba, Japan, 2014; Volume 54, p. 012001.

38. Du, S.; Li, H.; Xie, L.; Chen, L.; Peng, Y.; Zhu, Y.; Li, H.; Dong, P.; Wang, J. Vibrational frequencies of anti-diabetic drug studied by terahertz time-domain spectroscopy. *Appl. Phys. Lett.* **2012**, *100*, 143702. [CrossRef]

39. Ueno, Y.; Ajito, K.; Kukutsu, N.; Tamechika, E. Quantitative analysis of amino acids in dietary supplements using terahertz time-domain spectroscopy. *Anal. Sci.* **2011**, *27*, 351–351. [CrossRef]

40. Müller, J.; Brock, D.; Knop, K.; Zeitler, J.A.; Kleinebudde, P. Prediction of dissolution time and coating thickness of sustained release formulations using Raman spectroscopy and terahertz pulsed imaging. *Eur. J. Pharm. Biopharm.* **2012**, *80*, 690–697. [CrossRef]

41. Shen, Y.C. Terahertz pulsed spectroscopy and imaging for pharmaceutical applications: A review. *Int. J. Pharm.* **2011**, *417*, 48–60. [CrossRef]

42. Ergün, S.; Sönmez, S. Terahertz technology for military applications. *J. Manag. Inf. Sci.* **2015**, *3*, 13–16. [CrossRef]

43. Yang, X.; Zhao, X.; Yang, K.; Liu, Y.; Liu, Y.; Fu, W.; Luo, Y. Biomedical applications of terahertz spectroscopy and imaging. *Trends Biotechnol.* **2016**, *34*, 810–824. [CrossRef] [PubMed]

44. Sun, Q.; He, Y.; Liu, K.; Fan, S.; Parrott, E.P.; Pickwell-MacPherson, E. Recent advances in terahertz technology for biomedical applications. *Quant. Imaging Med. Surg.* **2017**, *7*, 345. [CrossRef] [PubMed]

45. Song, H.J.; Nagatsuma, T. *Handbook of Terahertz Technologies: Devices and Applications*; CRC Press: Roca Raton, FL, USA, 2015.

46. Saeedkia, D. *Handbook of Terahertz Technology For Imaging, Sensing And Communications*; Woodhead Publishing: Cambridge, UK 2013.

47. Dexheimer, S.L. *Terahertz Spectroscopy: Principles and Applications*; CRC Press: Boca Raton, FL, USA, 2007.

48. Sterczewski, L.A.; Nowak, K.; Szlachetko, B.; Grzelczak, M.P.; Szczesniak-Siega, B.; Plinska, S.; Malinka, W.; Plinski, E.F. Chemometric evaluation of THz spectral similarity for the selection of early drug candidates. *Sci. Rep.* **2017**, *7*, 14583. [CrossRef] [PubMed]

49. Byvatov, E.; Fechner, U.; Sadowski, J.; Schneider, G. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1882–1889. [CrossRef] [PubMed]

50. Korkmaz, S.; Zararsiz, G.; Goksuluk, D. Drug/nondrug classification using support vector machines with various feature selection strategies. *Comput. Methods Programs Biomed.* **2014**, *117*, 51–60. [CrossRef] [PubMed]

51. Burbidge, R.; Trotter, M.; Buxton, B.; Holden, S. Drug design by machine learning: Support vector machines for pharmaceutical data analysis. *Comput. Chem.* **2001**, *26*, 5–14. [CrossRef]

52. Louis, B.; Agrawal, V.K.; Khadikar, P.V. Prediction of intrinsic solubility of generic drugs using MLR, ANN and SVM analyses. *Eur. J. Med. Chem.* **2010**, *45*, 4018–4025. [CrossRef]

53. Sorich, M.J.; Miners, J.O.; McKinnon, R.A.; Winkler, D.A.; Burden, F.R.; Smith, P.A. Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2019–2024. [CrossRef]

54. Liu, M.; Wu, Y.; Chen, Y.; Sun, J.; Zhao, Z.; Chen, X.W.; Matheny, M.E.; Xu, H. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J. Am. Med Inform. Assoc.* **2012**, *19*, e28–e35. [CrossRef]

55. Heikamp, K.; Bajorath, J. Support vector machines for drug discovery. *Expert Opin. Drug Discov.* **2014**, *9*, 93–104. [CrossRef] [PubMed]

56. Chen, Y.; Liu, Y.; Zhao, G.; Wang, W.; Li, F. Chinese traditional medicine recognition by support vector machine (SVM) terahertz spectrum. *Guang Xue Guang Fen* **2009**, *29*, 2346–2350.

57. Nowak, M.; Nowak, K.; Grzelczak, M.; Szlachetko, B.; Sterczewski, L.; Plinski, E.; Swiatek, P.; Strzelecka, M.; Plinska, S.; Malinka, W. Machine learning applied to bi-heterocyclic drugs recognition. In Proceedings of the 2017 42nd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), Cancún, Mexico, 27 August–1 September 2017; pp. 1–2.

58. Selzer, P.M. *Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery*; Wiley-VCH: Weinheim, Germany, 2016.

59. Shibany, K.A.; Tötemeyer, S.; Pratt, S.L.; Paine, S.W. Equine hepatocytes: Isolation, cryopreservation, and applications to in vitro drug metabolism studies. *Pharmacol. Res. Perspect.* **2016**, *4*, e00268. [CrossRef] [PubMed]

60. Platt, J. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines; Technical Report MSR-TR-98-14, Microsoft Research. Available online: https://pdfs.semanticscholar.org/59ee/e096b49d66f39891eb88a6c84cc89acba12d.pdf (accessed on 28 June 2019).

61. Zdunek, R.; Nowak, M.; Plinski, E. Statistical classification of soft solder alloys by laser-induced breakdown spectroscopy: Review of methods. *J. Eur. Opt. Soc. Rapid Publ.* **2016**, *11*, 16006. [CrossRef]

62. Wojcik, M.R.; Zdunek, R.; Antonczak, A.J. Unsupervised verification of laser-induced breakdown spectroscopy dataset clustering. *Spectrochim. Acta Part At. Spectrosc.* **2016**, *126*, 84–92. [CrossRef]

63. Hofmann, T.; Schölkopf, B.; Smola, A. Kernel Methods in Machine Learning. *Ann. Stat.* **2008**, *36*, 1171–1220. [CrossRef]

64. Liu, W.; Principe, J.C.; Haykin, S. *Kernel Adaptive Filtering: A Comprehensive Introduction*, 1st ed.; Wiley Publishing: Hoboken, NJ, USA, 2010.

65. Gendreau, M.; Potvin, J.Y. *Handbook of Metaheuristics*, 2nd ed.; Springer Publishing Company, Incorporated: New York, NY, USA, 2010.

66. De Jong, K.A. *Evolutionary Computation: A Unified Approach*; MIT Press: Cambridge, MA, USA, 2016.

67. Lewis, R.M.; Torczon, V.; Trosset, M.W. Direct search methods: Then and now. *J. Comput. Appl. Math.* **2000**, *124*, 191–207. [CrossRef]

68. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [CrossRef]