

文章编号:1004-7220(2011)06-0555-05

基于时域太赫兹波谱法的降糖类药物分类与分析

李 晗¹, 董 萍¹, 王家泰¹, 杜少卿², 李 洪², 谢 乐², 陈 麟², 朱亦鸣^{2,3}

(1. 上海市杨浦区老年医院, 上海 200090; 2. 上海理工大学 教育部光学仪器与系统工程研究中心, 上海市现代光学系统重点实验室, 上海 200093; 3. 上海交通大学 微电子学院, 上海 200240)

摘要: 目的 根据降糖类药物在太赫兹频域范围的吸收光谱, 研究他们各自基团的振动频率及其共振峰, 以有效鉴别各种相似药品, 同时也为明确不同分子基团振荡在药理上的贡献打下基础。方法 利用时域太赫兹波谱系统对格列喹酮、格列吡嗪、格列齐特、格列美脲、瑞格列奈和二甲双胍 6 种降糖药进行检测, 得到他们在 0.3 ~ 3.0 THz 范围内的吸收谱; 同时选取 1.5 ~ 2.0 THz 范围内的分类特征数据, 用支持向量机的方法, 分别对这些降糖药的特征数据进行鉴别。结果 4 种磺脲类降糖药都在 1.37 THz 处有一明显的共振峰。根据吸收谱, 能够很轻易地区分出瑞格列奈和二甲双胍与磺脲类药物的差别; 而在支持向量机的帮助下, 4 类磺脲类药物也可被 100% 准确区分出来。结论 太赫兹作为一种新的检测手段, 在药物鉴别、药物品质控制、化学键/功能基团识别等方面有着积极的意义。
关键词: 太赫兹光谱; 支持向量机; 降糖药; 振动分析; 频段; 分子

中图分类号: R 318.01 文献标志码: A

Identification and analysis of anti-diabetic drugs based on terahertz time-domain spectroscopy

LI Han¹, DONG Ping¹, WANG Jia-tai¹, DU Shao-qing², LI Hong², XIE Le²,

CHEN Lin², ZHU Yi-ming^{2,3} (1. *Shanghai Yangpu Geriatric Hospital, Shanghai 200090, China;*

2. *Engineering Research Center of Optical Instrument and System, Ministry of Education, Shanghai Key Lab of Modern Optical System, University of Shanghai for Science and Technology, Shanghai 200093, China;*

3. *School of Microelectronics, Shanghai Jiaotong University, Shanghai 200240, China*)

Abstract: Objective To study the vibration frequencies and resonant peaks of different molecular groups in anti-diabetic drugs and to investigate the absorption spectra of these drugs in the range of terahertz, so as to accurately and efficiently identify similar drugs and provide foundation for understanding the contribution of such vibration in different molecular groups to pharmacology. Methods Using terahertz time-domain spectroscopy (THz-TDS), six kinds of similar diabetes pills for treating diabetes mellitus (DM), including gliquidone, glipizide, gliclazide, glimepiride, repaglinide and metformin were tested to obtain their absorption spectra from 0.3 ~ 3.0 THz. Support vector machines (SVM) method was employed to separate these anti-diabetic drugs by selecting the experimental data from 1.5 ~ 2.0 THz. Results For gliquidone, glipizide, gliclazide, and glimepiride, an obvious resonance peak was found at 1.37 THz. According to the THz spectra, repaglinide and metformin could be easily separated from sulfonylurea drugs. Furthermore, with the help of SVM, the accuracy of discrimination for four kinds of sulfonylurea could reach 100%. Conclusions THz technology, as a new testing method, shall play a positive role in the drugs for their identification, quality control and distinguishing their chemical bonds/functional group. Key words: Terahertz spectroscopy; Support vector machines; Anti-diabetic drugs; Vibration analysis; Frequency band; Molecules

收稿日期:2011-03-30; 修回日期:2011-05-05

基金项目:上海市科委纳米技术专项(0952nm02400),上海市科委浦江人才计划(09PJ1407800),国家教育部博士点新教师基金(20093120120007),教育部留学回国科研启动基金资助,国家自然科学基金资助项目(61007059)。

通讯作者:朱亦鸣,教授,Tel:(021)33773176;E-mail:ymzhu@usst.edu.cn。

太赫兹波(terahertz waves, $1 \text{ THz} = 10^{12} \text{ Hz}$)是介于毫米波和红外光之间的电磁辐射波,在无线电物理领域称为亚毫米波,在光学领域则习惯称之为远红外辐射波。通常所研究的 THz 辐射是指频率在 $0.1 \sim 10 \text{ THz}$, 波长在 $30 \mu\text{m} \sim 3 \text{ mm}$, 波数在 $3.3 \sim 330 \text{ cm}^{-1}$ 之间的电磁波。这一频率范围内的电磁波具有丰富的科学内容和广阔的应用前景。首先, THz 光子的能量只有毫电子伏特,与 X 射线相比,不会因为电离而破坏被测物质,从而可以安全地用于生物医学方面的检测和诊断;因此,THz 波在生物成像、医疗诊断等方面(如皮肤癌、烧伤诊断)具有广泛的应用。此外,THz 光谱在生物分子的结构和动力学特性等方面有着重要的研究价值。由于组成物质的分子各不相同,会产生不同的相互作用力(如范德瓦尔斯力等),从而使得每种物质基团振动的频率不同;而大部分生物分子的基团振动都落在 THz 波段,所以利用 THz 技术可以获得这些生物分子在 THz 波段的光学常数,进而研究他们在分子层面的特性。近些年来,已有国内外学者利用 THz 光谱对 DNA、蛋白质和糖类等生物分子进行了探索研究^[1-5]。

由于物质生活的丰富及不健康的生活方式,导致糖尿病成了危害人类健康的主要疾病之一。目前,针对 II 型糖尿病患者,临床上大多使用磺脲类口服降糖药进行治疗^[6]。迄今为止,磺脲类降糖药已有 3 代产品:甲苯磺丁脲、氯磺丙脲等是第一代;格列苯脲、格列吡嗪、格列齐特及格列喹酮等是第二代;格列美脲等为第三代。随着人们对其认识的加深,第一代产品因不良反应较大已经逐渐退出市场;第二代产品是在第一代药物的苯环上接一带芳香环的碳酰胺而成,作用是第一代药物的十几至上百倍,副作用较少发生,因而被广泛应用;第三代产品在磺酰脲的尿素部分加一个二环杂环,使该类药不仅具有降糖作用,而且能改善血小板功能,对糖尿病患者容易凝血和有血管栓塞倾向可能有益。针对瑞格列奈、二甲双胍和磺脲类等降糖药,因其分子结构上的差异,致使由于分子之间的力学特性所引起的不同基团振动模式得以在 THz 光谱中显现出来。

本文拟利用时域 THz 波谱技术对格列喹酮、格列吡嗪、格列齐特、格列美脲、瑞格列奈和二甲双胍 6 种降糖药进行检测,对采集到的时域信号作了傅

里叶变换,获取了 6 种样品的 THz 的频域谱和吸收谱。实验发现 4 种磺脲类降糖药都在 1.37 THz 处有一个明显的共振峰。根据吸收谱,能很轻易地区分出瑞格列奈和二甲双胍与磺脲类药物的差别。同时,结合支持向量机(SVM)理论^[7],通过对实验得到的 4 个磺脲类药物样品的 THz 指纹吸收谱与他们的基团振动关系进行分析,成功地区分出了这 4 种磺脲类药物,准确率达到了 100%,为 THz 技术应用于药物鉴别、药物品质控制、化学键/功能基团识别等方面提供了可行性。

1 THz 光谱实验

使用上海理工大学教育部光学仪器与系统工程研究中心下属超快光电子与太赫兹技术实验室的 THz 时域波谱检测系统(见图 1)。采用反射式产生 THz 脉冲的方法,所用的发射极为 GaAs 晶体,探测极为 ZnTe 晶体。实验中所使用的激光器是锁模钛宝石飞秒脉冲激光器。激光器出射功率为 1.3 W , 脉冲的中心波长为 800 nm , 脉冲宽度为 100 fs , 激光器重复率为 76 MHz 。飞秒脉冲激光被分束镜分成泵浦光和探测光,泵浦光照射到砷化镓样品上,激发出本征超纯砷化镓层中的电子和空穴对。电子和空穴对被电场加速,并向空间辐射出 THz 电磁波。

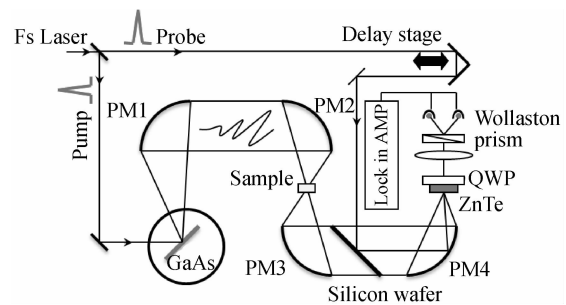


图 1 时域 THz 波谱检测系统

Fig. 1 Terahertz time-domain spectroscopy(THz-TDS) system

如图 1 所示,第 1 组离轴抛物面镜 PM1 和 PM2 将 THz 波聚集到被测样品上,第 2 组离轴抛物面镜 PM3 和 PM4 将携带有样品信息的 THz 波聚集在 $700 \mu\text{m}$ 的 $\langle 110 \rangle$ ZnTe 电光晶体上,探测光经过步进电机后由 Si 片反射也照射在电光晶体上。利用电光晶体的超快电光取样过程,测定出随时变化的 THz 时域信号^[8-11]。对于含有样品信息的 THz 脉冲

波形进行傅里叶变换, 得其 THz 频谱。通过含有样品信息的 THz 频谱和参考信号频谱的对比, 可以计算出样品的相对吸收系数。其公式为

$$a(\omega) = -\ln\left(\frac{I_{\text{sam}}(\omega)}{I_{\text{ref}}(\omega)}\right) / d \quad (1)$$

式中: d 为样品厚度, $I_{\text{sam}}(\omega)$ 是含有样品信息的透射谱, $I_{\text{ref}}(\omega)$ 是参考信号的透射谱 (THz 信号不经过样品, 直接被 ZnTe 电光晶体探测)。本实验中 6 种糖尿病治疗药物的样品厚度均控制在 0.6 mm (± 0.02 mm), 信噪比约为 1 000 : 1。通过加氮气的方法, 将实验湿度降低到 8% 以下, 以减少空气中的水蒸气对 THz 的吸收作用。实验中环境温度维持在 22.0 ~ 22.5 °C。由于抛物面镜直径大小的限制, 使得 0.3 THz 以下频率范围的 THz 信号无法被完整的收集到, 同时因受电光晶体厚度与飞秒激光脉宽的影响, THz 信号在频域上达到 3 THz 后截止, 故该检测系统有效频谱范围在 0.3 ~ 3.0 THz。另外, 对每一种样品都测试了 4 次, 以使实验数据更为准确。

2 实验数据处理与分析

为了便于观察, 先将所测的 4 组信号做了平均处理, 得到格列喹酮、格列吡嗪、格列齐特、格列美脲、瑞格列奈和二甲双胍各自的时域谱、频域谱和吸收谱。

图 2 所示为 6 种降糖药的时域波形。从形状上来看, 除了瑞格列奈在 1 ps 的位置上有一小峰外,

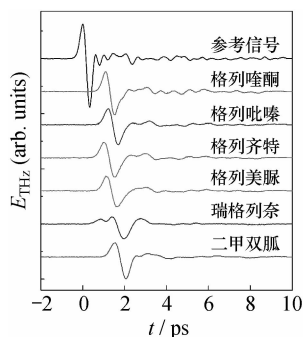


图 2 6 种降糖药的 THz 时域波形图

Fig. 2 Time-domain THz transmission waveforms for six kinds of anti-diabetic drugs

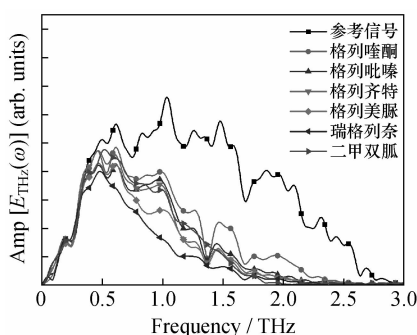


图 3 由 THz 时域信号傅立叶变换得到的 6 种降糖药的 THz 频域波形图

Fig. 3 THz transmission spectra for six kinds of anti-diabetic drugs obtained from the Fourier transformation of time-domain THz transmission traces

很难区分出其他药品之间的显著差异。图 3 所示为经过傅里叶变换后得到的含有样品信息的 THz 频谱。首先, 可以清楚地观察到, 在大于 2.5 THz 的高频段, 这 6 种降糖药的测试样品对 THz 电磁波有很强的吸收, 使得其频谱信号几乎为零。其次, 4 种磺脲类降糖药在 1.37 THz 的位置有着明显的下凹趋势, 而瑞格列奈与二甲双胍频谱曲线差异较大。图 4 所示为由式(1)所得这 6 种降糖药物的相对吸收系数, 从吸收谱上能更清晰的发现 1.37 THz 处存在着磺脲类药物的共振吸收峰, 并且 4 种磺脲类降糖药的吸收系数曲线十分接近, 不易分辨。此外, 二甲双胍在 1.86 THz 处有一强吸收峰, 瑞格列奈在 1.68 THz 和 1.87 THz 两处各有一个较明显的共振吸收峰。实验所得 6 种降糖药在本实验系统有效探测频谱范围内的不同吸收系数曲线是他们各自基团振动的真实体现, 而这些基团的振动模式又与药物分子结构息息相关; 因此, 借助于 THz 光谱还可以间接地了解药物内部化学键和官能团的组合方式, 在确定药物的功能和鉴定药物是否受到污染等方面发挥了一定的作用。

从频域信号 (见图 3) 和吸收系数 (见图 4) 曲线可知, 在 0.3 ~ 2.5 THz 的最有效频域范围内, 二甲双胍、瑞格列奈和 4 种磺脲类降糖药容易简单地被区分。然而对于 4 种磺脲类: 格列喹酮、格列吡嗪、格列齐特和格列美脲, 虽然都有明显的吸收峰, 但其所在位置都落在 1.37 THz, 无法简单地通过吸收谱进行区分。因此, 本研究又运用 SVM 加以区分。

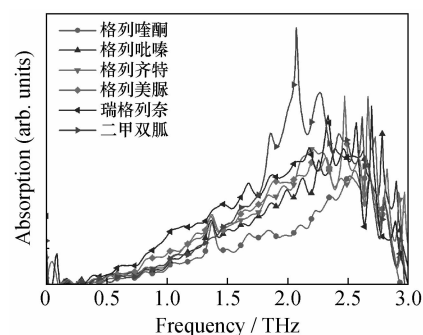


图 4 由 THz 透射谱得到的 6 种降糖药在 THz 波段的相对吸收系数

Fig. 4 Absorption spectra for six kinds of anti-diabetic drugs obtained from THz transmission spectra

SVM 是 20 世纪 90 年代中期出现的一种新的机器学习方法,能够较好地解决神经网络难以解决的小样本、非线性、高维数和存在局部极小点等实际问题,很快成为了继神经网络研究之后新的研究热点。目前,SVM 在食品和药品的检测和鉴定领域取得了很多应用成果^[12]。鉴于此,本文将 SVM 与磺脲类降糖药的太赫兹光谱技术相结合,对同种类的降糖药进行鉴别。

SVM 方法^[13-14]是根据有限的样本信息在模型的复杂性(即对特定训练样本的学习精度)和学习能力(即无错误地识别任意样本的能力)之间寻求最佳折衷,以期获得最好的推广能力。

本文中,选用了“1 对 1”的方法^[15]对样品进行多分类预测。该方法对于 k 个类,需要建立 $k(k-1)/2$ 个分类器,每次选一个类作为正类,选一个类作为负类,进行二分类,其最终结果由多个二分类叠加而成。为了比较全面地了解核函数对分类效果的影响,实验采用 3 种最常用的核函数,并且使用相同的核函数参数。函数形式为:

线性内积函数

$$k(x, y) = x \cdot y \quad (2)$$

多项式内积函数

$$k(x, y) = [(x \cdot y) + 1]^d \quad (3)$$

径向基内积函数

$$k(x, y) = \exp\{-|x - y|^2/\delta^2\} \quad (4)$$

二分类 SVM 的参数主要有 c 和 g 2 个。 c 是二分类 SVM 的惩罚因子,作用是在确定的特征空间中调节学习机器的置信范围和经验风险的比例以调节其推广能力。 g 就是式(4)中的 δ^2 ,当 δ^2 太小时,会产生对测试样本不具有任何泛化能力的严重“过学习”现象。而当 δ^2 太大时,则会产生把所有样本都划分为一类的“欠学习”现象。

如图 5 所示,选取了 4 种样品在 1.5 ~ 2.0 THz 范围内的吸收系数数据作为识别的特征信息,观察它们在不同频率点上吸收强度值的分布情况。由于实验仪器本身的噪声等原因,对相同样品检测 4 次后得到的数据会存在一定误差,其结果就导致了 4 种磺脲类降糖药的相对吸收系数曲线都在一个小范围内有所波动。将所有 4 组数据中的 3 组作为样本集分别输入用 3 种核函数构建的 SVM 进行训练。在 SVM 的学习下,算法将生成能够区分这 4 种药物

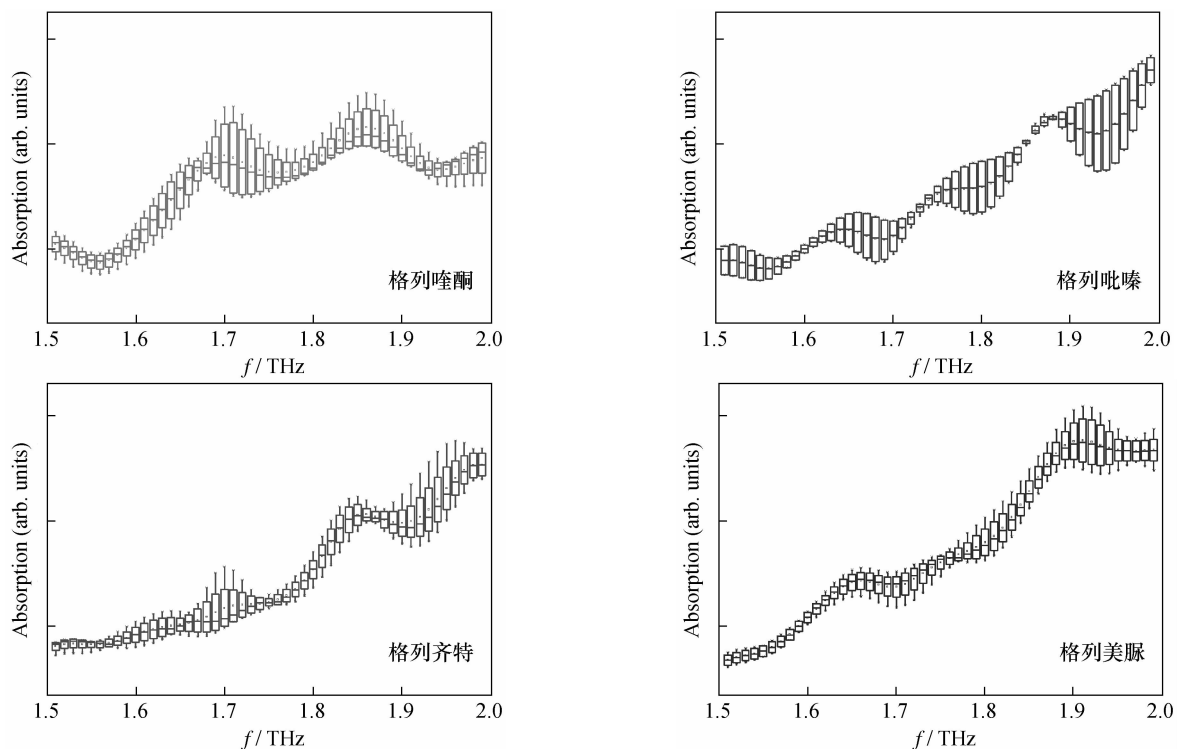


图 5 4 类磺脲类降糖药在 1.5 ~ 2.0 THz 频域内相对吸收系数分布图

Fig. 5 Distribution of absorption spectra (1.5 ~ 2.0 THz) for four kinds of sulfonylurea drugs

的分类面,紧接着预测出每一类药物所在的特定区间,若新的未被分类的数据落在其中的某个区间,即可判定其药物的类别。

本实验中,总共测量了4组数据(4种样品,共16个数据)。任意选择其中的3组数据(4种样品,共12个数据),并假定是已知的物质,作为样本集分别输入SVM进行训练。剩余的1组数据(4种样品,共4个数据)假定是未知的物质,被作为验证数据输入SVM中,以验证分类的准确性。

在3种常用核函数模型($c=1, g=0.25$)下,所用SVM算法对于4种磺脲类降糖药的区分能力均达到100%正确率的结果(见表1)。

表1 运用3种常用核函数进行SVM预测的结果

Tab.1 Results of classification with three SVM analyses

核函数类型	预测正确率/%
线性内积函数	100
多项式内积函数	100
径向基内积函数	100

3 结论

本文利用时域THz波谱检测技术对格列喹酮、格列吡嗪、格列齐特、格列美脲、瑞格列奈和二甲双胍6种降糖药进行太赫兹波段的频谱分析。实验发现在1.37 THz的位置上,格列喹酮、格列吡嗪、格列齐特和格列美脲这4种磺脲类降糖药都有一明显的共振峰。同时根据所得到的含有样品信息的指纹吸收谱,能够简单地区分出瑞格列奈、二甲双胍和4种磺脲类药物。最后结合SVM理论,在1.5~2.0 THz范围内对4种磺脲类样品作进一步的分析。在所选择的3种核函数(线性内积函数、多项式内积函数和径向基内积函数)下的SVM识别准确率均达到100%。

每一种大分子基团在THz频率范围内都有其独一无二的振动模式,这些振动模式又与吸收光谱一一对应。因此,通过对THz光谱的研究分析,可以获取药物中许多官能团的信息,以此确定部分乃至全部分子类型及结构,这在药物鉴别、药物品质控制、化学键/功能基团识别等方面都有着广阔的发展趋势。至于THz光谱是否与药物疗效有直接或间接的关联,还需进一步的探索。

参考文献:

- [1] Liu HB, George P, Sarah E, *et al.* Sensing minute changes in biological cell monolayers with THz differential time-domain spectroscopy [J]. *Biosens Bioelectron*, 2007, 22(6): 1075-1080.
- [2] Jepsen PU, Moller U, Merbold H. Investigation of aqueous alcohol and sugar solutions with reflection terahertz time-domain spectroscopy [J]. *Opt Express*, 2007, 15(22): 14717-14737.
- [3] Melinger JS, Laman N, Grischkowsky D. The underlying terahertz spectrum of explosives solids [J]. *Appl Phys Lett*, 2008, 93(011102): 1-3.
- [4] Liu GF, Ma XJ, Ma SH, *et al.* Identification of explosives 2,4-DNT and 2,6-DNT using terahertz time-domain spectroscopy [J]. *Chin J Chem*, 2008, 26(7): 1257-1261.
- [5] Ma SH, Liu GF, Zhang P, *et al.* The THz fingerprint spectra of the active ingredients of a TCM-Herba Ephedrae [J]. *Proc SPIE*, 2009, 7277(727711): 1-10.
- [6] Janbon M, Chaptal J, Vedel A, *et al.* Accidents hypoglycémiques graves par un sulfamidothiazol (le VK 57 ou 2254 RP) [J]. *Montp Med*, 1942, 441: 21-22.
- [7] Cortes C, Vapnik V. Support-vector networks [J]. *Mach Learn*, 1995, 20: 273-297.
- [8] Zhu YM, Unuma T, Shibata K, *et al.* Power dissipation spectra and terahertz intervalley transfer gain in bulk GaAs under high electric fields [J]. *Appl Phys Lett*, 2008, 93(232102): 1-3.
- [9] Zhu YM, Unuma T, Shibata K, *et al.* Femtosecond acceleration of electrons under high electric fields in bulk GaAs investigated by time-domain terahertz spectroscopy [J]. *Appl Phys Lett*, 2008, 93(042116): 1-3.
- [10] Wu Q, Zhang XC. Free-space electro-optics sampling of mid-infrared pulses [J]. *Appl Phys Lett*, 1997, 71(10): 1-3.
- [11] Wu Q, Zhang XC. 7 Terahertz broadband gap electro-optic sensor [J]. *Appl Phys Lett*, 1997, 70(14): 1-3.
- [12] 陈全胜, 赵杰文, 张海东, 等. 基于支持向量机的近红外光谱鉴别茶叶的真伪[J]. *光学学报*, 2006, 26(6): 933-937.
- [13] 边肇祺, 张学工. 模式识别 [M]. 北京:清华大学出版社, 2000: 284-299.
- [14] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods [M]. Cambridge: Cambridge University Press, 2000: 26-51.
- [15] Knerr S, Personnaz L, Dreyfus G. Single-layer learning revisited: A stepwise procedure for building and training a neural network. In Fogelman J ed. *Neurocomputing: algorithms, architectures and applications* [M]. Berlin: Springer-Verlag, 1990: 41-50.