Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 170 (2017) 117–123

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Firefly algorithm versus genetic algorithm as powerful variable selection tools and their effect on different multivariate calibration models in spectroscopy: A comparative study

Khalid A.M. Attia [a], Mohammed W.I. Nassar [a], Mohamed B. El-Zeiny [b], Ahmed Serag [a,*]

[a] Pharmaceutical Analytical Chemistry Department, Faculty of Pharmacy, Al-Azhar University, 11751 Nasr City, Cairo, Egypt
[b] Analytical Chemistry Department, Faculty of Pharmacy, Modern University for Technology and Information (MTI), 12582 Al Hadaba Al Wosta, Cairo, Egypt

ARTICLE INFO

Article history:
Received 27 April 2016
Received in revised form 2 July 2016
Accepted 8 July 2016
Available online 10 July 2016

Keywords:
Firefly algorithm
Genetic algorithm
Concentration residual augmented classical least squares
Artificial neural network
Support vector regression

ABSTRACT

For the first time, a new variable selection method based on swarm intelligence namely firefly algorithm is coupled with three different multivariate calibration models namely, concentration residual augmented classical least squares, artificial neural network and support vector regression in UV spectral data. A comparative study between the firefly algorithm and the well-known genetic algorithm was developed. The discussion revealed the superiority of using this new powerful algorithm over the well-known genetic algorithm. Moreover, different statistical tests were performed and no significant differences were found between all the models regarding their predictabilities. This ensures that simpler and faster models were obtained without any deterioration of the quality of the calibration.

1. Introduction

Multivariate calibration has been widely applied in many fields, such as pharmaceutical, medical and environmental research to develop quantitative relations between variables and properties of interest [1]. Large number of variables compared with the number of samples is considered one of the greatest challenges in multivariate calibration [2]. So, selecting the most informative variables or eliminating the uninformative ones can still improve the performance of multivariate calibration models. Hence, variable selection techniques are considered one of the most promising areas of research in chemometrics. Greater efforts are directed to variable selection techniques based on swarm intelligence and nature inspired algorithms. The swarm intelligence algorithms are based on the interaction between agents from the same population, as well as, on the interaction with the environment which is their major advantage over other algorithms. So, they use crowd decision rather than random search. These agent-based algorithms are normally nature-inspired, e.g., the source of inspiration being ants' colonies [3], or flock of birds behavior [4].

For the first time, firefly as a variable selection algorithm [5–8] in UV spectral data was introduced in combination with three different multivariate models namely, concentration residual augmented classical least squares (CRACLS) [9,10], artificial neural network [11–13] (ANN) and support vector regression (SVR) [14–18]. Also, a comparative study was developed between this algorithm and the well-known genetic algorithm [19–21] on the same multivariate models. This study was applied for the determination of ciprofloxacin (CIP) Fig. 1 **(a)** in the presence of metronidazole (MET) as interferent Fig. 1 **(b)** in laboratory prepared mixtures and in their pharmaceutical dosage form.

2. Experimental

2.1. Materials and reagents

A. Pure CIP (certified to contain 99.25%) and MET (certified to contain 99.65%) were kindly supplied by Minapharm Pharmaceutical Company, Cairo, Egypt.
B. Ciprodiazole tablets nominally containing CIP (500 mg) and MET (500 mg) batch number EJE3135 were manufactured and supplied by MINAPHARM pharmaceuticals (Cairo, Egypt).
C. Methanol; El-NASR Pharmaceutical Chemicals Co., Egypt.

2.2. Instruments

SHIMADZU dual beam UV–visible spectrophotometer (Kyoto/Japan), model UV-1800 PC connected to IBM compatible and a

* Corresponding author.
E-mail address: Ahmedserag777@hotmail.com (A. Serag).

http://dx.doi.org/10.1016/j.saa.2016.07.016
1386-1425/

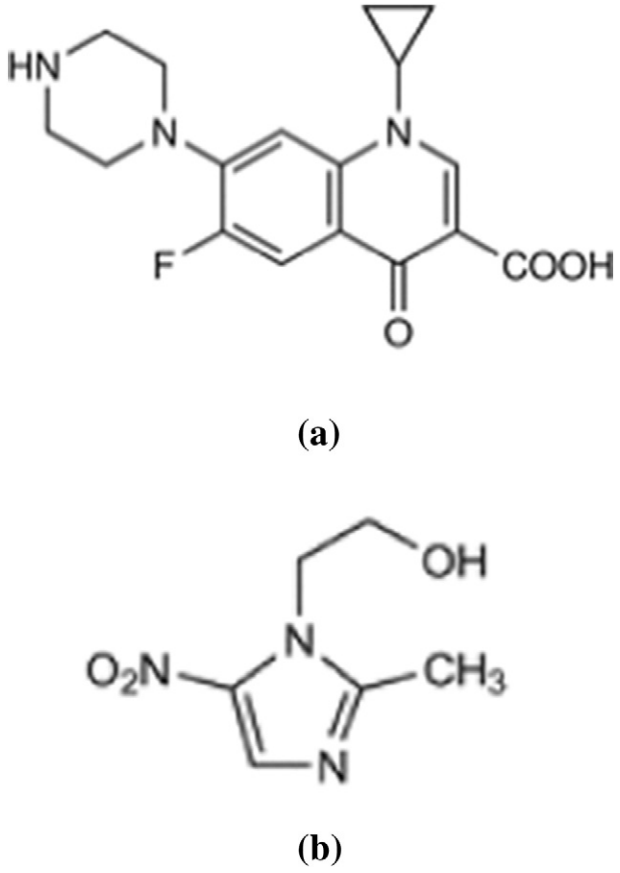

Fig. 1. Chemical structure of (a) ciprofloxacin and (b) metronidazole.

HP1020 laser jet printer. The spectral band was 2 nm and scanning speed is 2800 nm/min with 1 nm interval.

2.3. Software

The bundled software, UV-Probe personal spectroscopy software version 2.43 (SHIMADZU) was used. All chemometric methods were implemented in Matlab 8.2.0.701 (R2013b). Grid search for optimum SVR parameters, CRACLS and FA were done with our own written codes in Matlab. The codes for the SVR algorithm were downloaded from the internet website http://onlinesvr.altervista.org/ [22]. ANNs were carried out by using Neural Network toolbox. The *t*-test and *F*-test were performed using Microsoft® Excel. One way ANOVA test was performed using Graph Pad Prism version 5 (Graph Pad, San Diego, CA).

3. Procedures

3.1. Standard solutions

A. CIP and MET standard stock solutions prepared to contain 200 μg mL^{-1} in methanol.
B. CIP and MET standard working solutions prepared to contain 50 μg mL^{-1} in methanol.

3.2. Experimental design for chemometric models

A 5-level, 2-factor design was performed using 5 concentration levels for each of the 2 compounds resulting in 25 mixtures [23]. The central level of the design is 5 μg mL^{-1} for each of CIP and MET. The chosen concentrations for each compound is based on their linearity and the ratio between the two compounds involved in their pharmaceutical preparation. Table 1 represents the concentration design matrix while Fig. 2 shows the absorption spectra of these concentrations. Since, calibration data set selection based on optimal criteria improves the quality of multivariate models predictions, D-optimal selection algorithm was used for the calibration data set selection [24]. This algorithm improves the representativeness of the calibration data as shown by the 2D plot of the experimental space in Fig. 3. This plot shows the positioning of training set and the validation set samples.

Fifteen mixtures of this design were used as a calibration set and the other ten mixtures were used as a validation set to test the predictability of the developed multivariate models.

3.3. Analysis of CIP in Ciprodiazole® tablets by the proposed methods

Ten **Ciprodiazole®** tablets were weighed, powdered and mixed. The appropriate weight of powder equivalent to 20 mg of CIP was accurately transferred to 100-mL volumetric flask and the volume was made up to 100 mL with methanol. The solution was shaken vigorously for 15 min then sonicated for 30 min and filtered. Working solution was obtained by dilution of the stock solution with methanol to get solution labeled to contain (50 μg mL^{-1}). Necessary dilutions were made with methanol to obtain the different concentrations of the studied drug. The spectra of these solutions were scanned, stored in the computer and analyzed by the proposed models.

4. Results and discussion

4.1. Variable selection by FA and GA

In order to increase the quality of the calibration, FA as a variable selection tool has been introduced here for the first time with three multivariate calibration models to solve the severe overlapping of CIP and MET as shown in Fig. 4. This powerful variable selection tool has been compared with the most widely used GA listing the advantages and the disadvantages of each method.

FA was run on the calibration data to determine the selected variables using RMSE as a fitness function calculated by MLR model. It

Table 1
The concentrations of CIP and MET in μg mL^{-1} in the used experimental design.

Mixture number	CIP	MET
1	5	5
2	5	4
3	4	4
4	4	6
5	6	4.5
6	4.5	6
7	6	5
8	5	4.5
9	4.5	4.5
10	4.5	5.5
11	5.5	6
12	6	5.5
13	5.5	5
14	5	6
15	6	6
16	6	4
17	4	5.5
18	5.5	4
19	4	5
20	5	5.5
21	5.5	5.5
22	5.5	4.5
23	4.5	4
24	4	4.5
25	4.5	5

The shaded rows represent the calibration set.