ISSN: 2357-0547 (Print)
ISSN: 2357-0539 (Online)

Research Article / JAPR
Salama et al., 2017, 1 (4), 185-192

Journal of Advanced Pharmacy Research

Multivariate Chemometric Models and Application of Genetic Algorithm for Simultaneous Determination of Ledipasvir and Sofosbuvir in Pure Form and in Pharmaceutical Preparation; A Comparative Study

Fathy M. Salama, Khalid A. Attia, Ahmed A. Abouserie, Ahmed El-Olemy, Ebrahim Abolmagd*

Pharmaceutical Analytical Chemistry Department, Faculty of Pharmacy, Al-Azhar University, 11751, Nasr City, Cairo, Egypt

***Corresponding author*:** Ebrahim Abolmagd, *Pharmaceutical Analytical Chemistry Department, Faculty of Pharmacy, Al-Azhar University, 11751, Nasr City, Cairo, Egypt. Tel.: +201284955570*
E-mail address: Eb.Abolmagd@yahoo.com

Submitted on: 29-03-2017; Revised on: 04-07-2017; Accepted on: 12-07-2017

ABSTRACT

Objectives: Four multivariate chemometric methods have been developed for simultaneous determination of sofosbuvir and ledipasvir in their pure and pharmaceutical dosage forms. **Methods**: Firstly, partial least squares and artificial neural network have been applied for the quantitative analysis of the studied drugs. **Results**: Experimental design of different synthetic mixtures of sofosbuvir and ledipasvir in different ratios has been done. The zero-order absorption spectra of these prepared mixtures have been recorded over the wavelength range 200-400 nm with 1 nm interval. The obtained absorbance and concentration data matrix have been utilized to obtain calibration or regression analysis data which has been used for the prediction of the unknown concentrations of each drug in their mixtures. Alternatively, the application of genetic algorithm to partial least squares and artificial neural network has been done and greatly increased the precision and predictive ability of the methods. The four methods have been successfully applied to quantify sofosbuvir and ledipasvir in the real market sample. **Conclusion:** The investigated methods have been found to be accurate, precise and could resolve the overlapped spectra of the mixture without any preliminary separation steps.

Keywords: *Chemometry; Ledipasvir; Overlapped spectra; Sofosbuvir.*

INTRODUCTION

The new anti-viral combination *Sofolanork plus*® containing Ledipasvir and sofosbuvir is one of the most effective protocol that can manage and completely cure the Hepatitis C virus patients.

Ledipasvir (LED); (**Figure 1**) is methyl *N*-[(2*S*)-1-[(6*S*)-6-[5-[9,9-difluoro-7-[2[(1*S*,2*S*,4*R*)3[(2*S*)-2(methoxy carbonyl amino)-3-methyl butanoyl]-3-azabicyclo[2.2.1]heptan-2-yl]-3H-benzimidazol-5-yl] fluoren-2-yl]-1H-imidazol-2-yl]-5-azaspiro[2.4] heptan-5-yl]-3-methyl-1-oxobutan-2-yl]carbamate. It is an inhibitor of an important viral phosphor-protein, NS5A, which is involved in viral replication, assembly and secretion.[1] Physically it is a white to tinted (off-white, tan, yellow, orange, or pink), slightly hygroscopic crystalline solid. It is freely soluble in methanol, ethanol and DMSO and slightly soluble in acetone. Its molecular weight is 889.[2]

Sofosbuvir (SOF); (**Figure 2**) is (*S*) - isopropyl 2-((*S*)-(((2*R*,3*R*,4*R*,5*R*)-5-(2,4dioxo- 3,4dihydro pyrimidin-1(2H)-yl)-4-fluoro-3 hydroxy-4 methyl tetra hydrofuran-2-yl)methoxy)-(phenoxy) phosphorylamino) propanoate. It is potent in inhibiting the HCV NS5B RNA-dependent RNA polymerase, it undergoes intracellular metabolism to produce GS-461203, active uridine analog triphosphate which inhibits the polymerase activity of the NS5B from HCV genotype 1b, 2a, 3a and 4a with IC_{50} values ranging from 0.7 to 2.6 µM estimated in a biochemical assay. Physically it is a white crystalline solid soluble in the pH range of 2-7.7 at temperature 37 ºC and has slight aqueous solubility. Its molecular weight is 529.45.[3]

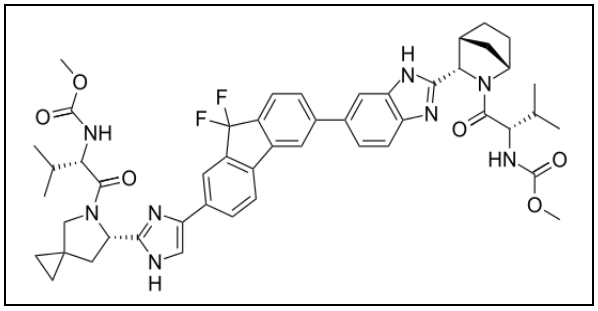

Figure 1. Structural formula of Ledipasvir.

Since this anti-viral combination is newly formulated, few analytical methods have been reported in the literature for simultaneous determination of SOF and LED. There is only two reversed-phase high-performance liquid chromatography methods[3,4] and two UPLC-MS/MS methods[5,6] available for the simultaneous determination of SOF and LED in dosage form, only few spectrophotometric methods were available for the simultaneous determination of LED and SOF in their dosage form[7,8].

To the best of our knowledge there is no chemometric methods available for simultaneous determination of ledipasvir and sofosbuvir.

Hence, the aim of this work was to develop an accurate and precise chemometric methods for simultaneous determination of LED and SOF in their dosage form. The developed methods are partial least squares (PLS-1) with application of genetic algorithm (GA-PLS-1) and artificial neural network (ANN) with application of genetic algorithm (GA-ANN).

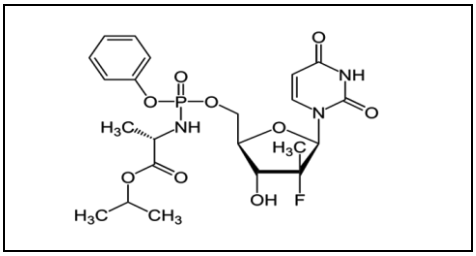

Figure 2. Structural formula of Sofosbuvir

MATERIALS AND METHODS

Instruments

Shimadzu UV-Vis. 1800 *Spectrophotometer*, (Tokyo, Japan), equipped with 10 mm matched quartz cells was used. The spectral band was 2 nm and scanning speed is 2800 nm/min with 1 nm interval.

Software

UV-Probe personal spectroscopy software version 2.1. (SHIMADZU).

All chemometric methods were implemented in Matlab R2013b (8.2.0.701).

PLS, ANN and application of GA were carried out by using PLS toolbox software version 2.1. in conjugation with neural network toolbox

The student *t*-test and F value were performed using Microsoft-Excel.

All calculations were performed using a Quad core CPU, 1.47 GHz, 4.00 GB of RAM under Microsoft Windows 7 ™.

Materials and Reagents

Pure LED and SOF were kindly supplied by Mash Premiere for Pharmaceutical and Cosmetics Industries, Third Industrial Zone, Badr City, Egypt. Their purity were (99.25 %) and (99.7 %) respectively according to the company certificates.

Pharmaceutical preparation: Sofolanork Plus® tablets (Batch no. M 169916), manufactured by Mash Premiere for Pharmaceutical and Cosmetics Industries. It is labelled to contain (400 mg of SOF and 90 mg of LED) per tablet and purchased from local pharmacies.

Methanol, analytical grade was purchased from (El-Nasr Pharmaceutical Chemicals Co. Abu-Zabaal, Cairo, Egypt).

Standard solutions

A *Standard solution* of LED (450 µg mL^{-1}) was prepared by dissolving 45 mg of LED in 50 mL of methanol and the volume was completed to 100 mL with methanol. *Working solution* of LED (45 µg mL^{-1}) was prepared by transferring 10 mL of standard solution to 100 mL volumetric flask and the volume was completed to 100 mL with methanol.

A *Standard solution* of SOF (1 mg mL^{-1}) was prepared by dissolving 100 mg of SOF in 50 mL of methanol and the volume was completed to 100 mL with methanol. *Working solution* of SOF (200 µg mL^{-1}) was prepared by transferring 20 mL of its standard solution to 100 mL volumetric flask and the volume was completed to 100 mL with methanol.

Methods

Experimental design

A 5 levels, 2 factors experimental design was used in which 0.8, 0.9, 1, 1.1 or 1.2 mL aliquots of both LED and SOF working solutions equivalent to (36, 40.5, 45, 49.5 and 54 µg ml^{-1}) for LED and (160, 180, 200, 220 and 240 µg ml^{-1}) of SOF were combined and diluted to 10 mL with methanol resulting in 25 mixtures.[9] The central level of the design is 4.5 µg ml^{-1} and 20 µg ml^{-1} for LED and SOF respectively. The chosen concentrations for each compound are based on their linearity and the ratio between both compounds involved in their pharmaceutical preparation. The concentrations details are given in **Table 1.**

The absorption spectra of the prepared mixtures were recorded over the wavelength range 200-400 nm with 1 nm interval thus the produced spectral data matrix has 25 rows representing different samples and 201 columns representing wavelengths with dimensions (25 x 201).

12 mixtures of this design (odd numbers) were used for the calibration based on cross validation and the other 12 mixtures (even numbers) were used as a validation set to test the predictability of the proposed multivariate models.

Application of the method to pharmaceutical formulation:

Ten tablets of Sofolanork Plus® (400/90 mg) was finely powdered and an amount equivalent to one tablet (400 mg of SOF and 90 mg of LED) was extracted three times with 25 mL of methanol, filtered into 100 mL volumetric flask then the volume was adjusted with methanol to obtain a solution labelled to contain (4000 µg mL^{-1} of SOF and 900 µg mL^{-1} of LED). This solution was diluted to obtain solution labelled to contain (400 µg mL^{-1} of SOF and 90 µg mL^{-1} of LED). The spectra of these solutions were scanned from 200 to 400 nm, stored in the computer and analysed by the proposed methods.

RESULTS AND DISCUSSION

Spectroscopic techniques can supply the analyst with a large data within a short period of time. Coupling the spectral data with chemometric models enhance the quality of the spectral information and making this combined technique into a powerful and highly convenient analytical tool. Few spectrophotometric methods have been introduced for the simultaneous analysis of SOF and LED. All these methods require manipulation either through derivatization of the absorption spectra, processing of the ratio spectra or requiring the presence of a critical point (iso absorptive point). This has prompted the authors to apply different chemometric methods, especially PLS, GA-PLS, ANN, GA-ANN for simultaneous analysis of the studied drugs. These described methods have higher prediction power, providing maximum relevant information and analyzing a large number of samples in a short period of time with higher degree of accuracy and precision.

The UV spectra of SOF and LED show certain degree of overlap **Figure 3**, which creates difficulty in the simultaneous analysis of this mixture. Therefore, multivariate calibration methods were applied to predict the concentrations of SOF and LED in both calibration and validation sets as well as in their pharmaceutical formulation.

GA searches the solution space of a function through the use of simulated evolution. It solves the optimization problem by exploring all regions of the potential solutions and exponentially exploiting promising areas through mutation, crossover, and selection operation applied to individuals in the populations. A critical issue of successful GA performance is the adjustment of GA parameters [10]. In order to avoid the risk of over fitting, a number of independent short runs was done and the results of all the runs were taken into consideration to obtain the final model. Doing this, a much more consistent (and less over fitted) solution can be obtained[11, 12]. The adjusted GA parameters with the lowest mean square error were shown in **Table 2.**

Table 1. The 5-level, 2-factor experimental design shown as concentrations of the mixture components in µg mL^{-1}.

Mixture number	LED	SOF
1	4.5	20
2	4.5	16
3	3.6	16
4	3.6	24
5	5.4	18
6	4.05	24
7	5.4	20
8	4.5	18
9	4.05	18
10	4.05	22
11	4.95	24
12	5.4	22
13	4.95	20
14	4.5	24
15	5.4	24
16	5.4	16
17	3.6	22
18	4.95	16
19	3.6	20
20	4.5	22
21	4.95	22
22	4.95	18
23	4.05	16
24	3.6	18
25	4.05	20

The shaded rows represent the validation set

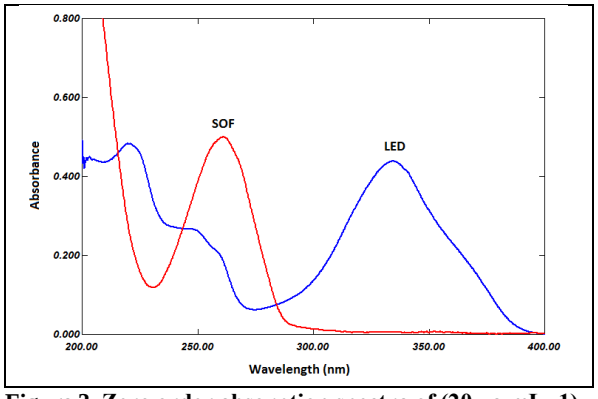

Figure 3. Zero order absorption spectra of (20 µg mL−1) SOF and (4.5 µg mL−1) LED

Partial least squares (PLS) and applying genetic algorithm (GA-PLS)

PLS-1 is a widely used regression method. It is known that information from the concentrations values is introduced into the calculation of the so-called latent variables, which are linear combinations of the original variables. PLS-1 method was run on the calibration data of absorption spectra. To select the number of factors in the PLS-1 algorithm, a cross validation (CV) method leaving out one sample at a time was applied using calibration set of 13 calibration spectra. RMSECV (Root Mean Squared Error of Cross Validation) was recalculated upon addition of each new factor to the PLS-1. Then number of factors was selected based on Haaland and Tomas criteria[13]. It was found that two factors were sufficient for modelling both SOF and LED.

Table 2. Parameters of Genetic algorithm.

Parameter	Value	
	LED	SOF
Population size	40	40
Maximum generations	42	48
Mutation rate	0.005	0.005
The number of variables in a window (window width)	3	2
Per cent of population the same at Convergence	100	100
% Wavelengths used at initiation	50	50
Crossover type	Double	Double
Maximum number of latent variables	2	2
Cross validation	Random	Random
Number of subsets to divide data into for cross validation	5	5
Number of iterations for cross validation at each generation	2	2

However, to increase the quality and improve the calibration, the variables selection technique namely genetic algorithm (GA) was performed, by its application the un-informative variables were excluded. The predictability of both models was tested by validation set and it was found that the PLS-1 model constructed after removing the un-informative variables is more robust and simpler with lower root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP). This is surely due to the fact that the un-informative wavelengths have been excluded. The percentage % recoveries, RSD (relative standard deviation) and RMSEP values of the validation set for PLS and GA-PLS models are listed in **Table 3**.

The GA was run on 201 variables for SOF and LED using a PLS with the optimum number of LVs determined by cross validation on the model containing all the variables. GA reduced absorbance matrix to about 49 % of the original matrix of LED and 45 % of SOF. The whole parameters involved upon application of GA on PLS model are shown in **Figures 4** and **5** for SOF and LED respectively.

Artificial neural network (ANN) and applying genetic algorithm (GA-ANN)

ANNs are a type of computational models simulating the biological neural networks. They composed of an inter-connected group of artificial neurons. To optimize a neural network, we have to use the trial and error method to find out the best neural network architecture[14,15]. Choosing the values of optimum parameters to construct the network is not an easy task because the parameters are mutually related.

The output layer resemble the concentration vector of one component. The hidden layer consists of single layer which is sufficient to solve similar or more complex problems. Moreover, more hidden layers may cause over-fitting. The hidden neurons number is one of the most important parameters among other ANN parameters that must be adjusted. This parameter is related to the converging performance of the output error function during the learning process.

Transfer function pair also an important parameter that should be adjusted carefully. Choosing of transfer function based on the nature of data to be analysed. In the present work, purelin-purelin transfer function was used due to the linear correlation between absorbance and concentration. The learning rate controls the degree at which connection weights are modified during the learning phase. The optimized parameters values of the ANN for SOF and LED were shown in **Table 4.**

ANNs show better RMSEP than PLS-1 which may be due to the fact that ANNs is a type of artificial intelligence where there is less chance for over-fitting than that may occur in PLS calibrations. % recoveries, % RSD and RMSEP values of the validation set for ANN and GA-ANN models are listed in **Table 3**.

The application of the ANN on the raw data after using the variable selection technique GA show improvement of the results. A large number of nodes in the input layer of the network (wavelengths) increases the CPU time for ANN modelling. GA allowed the use of less number of neurons (shorter training time) than those used in the network utilized the raw data

Analysis of real market sample

The proposed procedure was applied for determination of LED in presence of SOF in Sofolanork plus® tablets. Satisfactory results were obtained in good

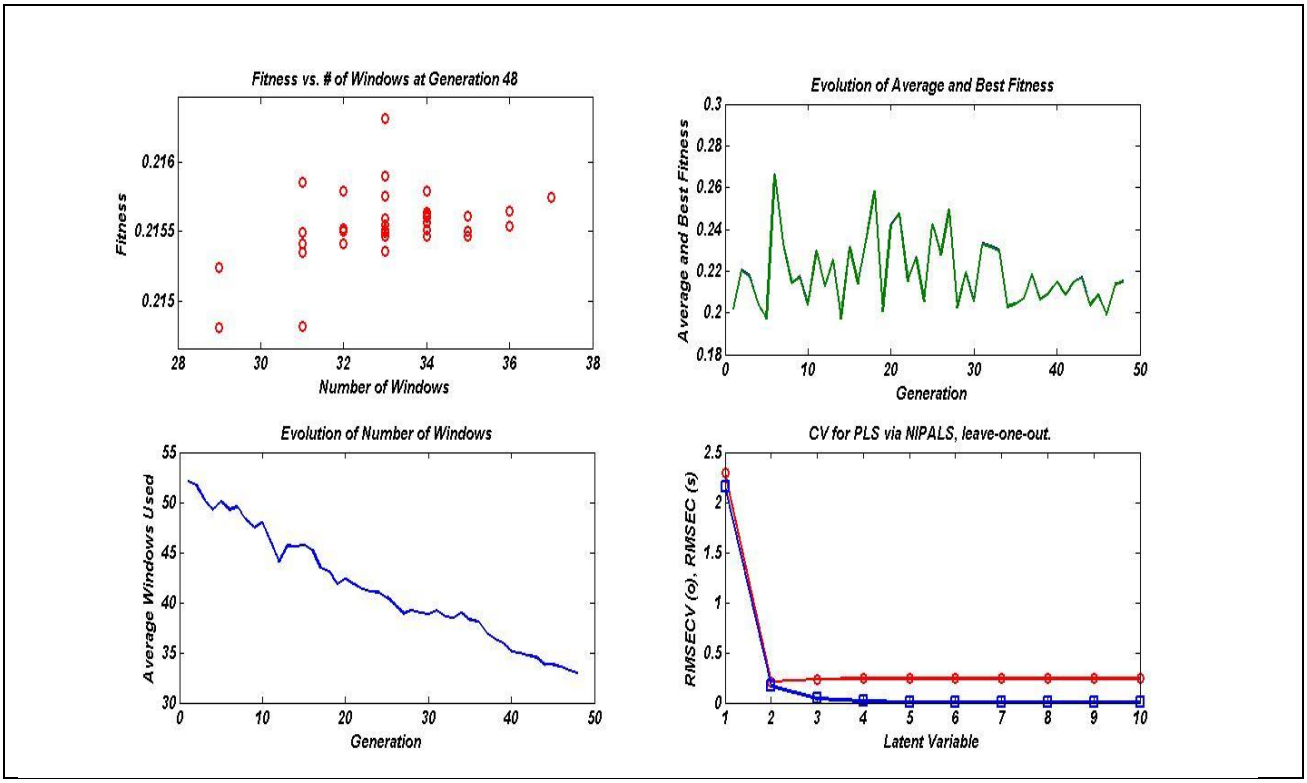

Figure 4. The whole parameters involved in application of GA on PLS model for SOF.

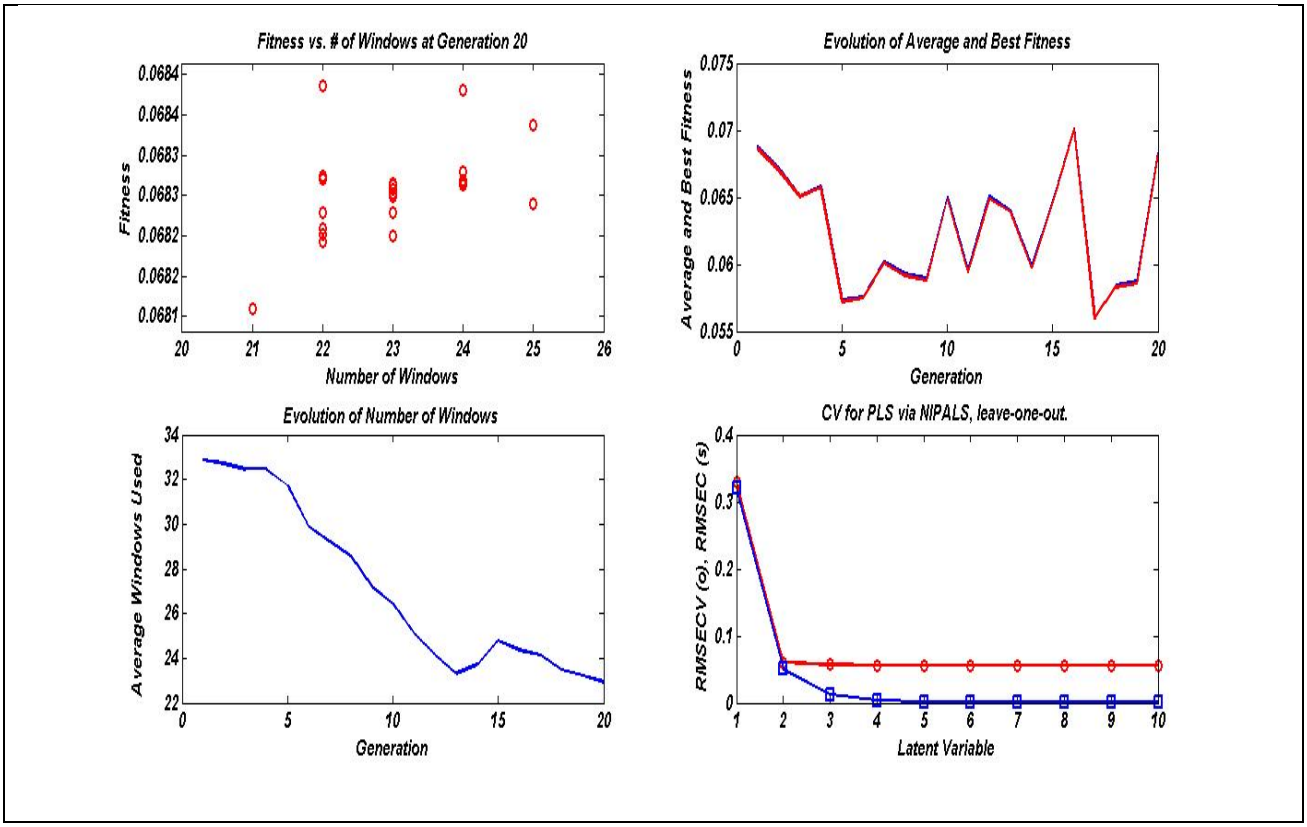

Figure 5. The whole parameters involved in application of GA on PLS model for LED

Table 3. Validation parameters of the proposed methods.

Mixture number	PLS		GA-PLS		ANN		GA-ANN	
	LED	SOF	LED	SOF	LED	SOF	LED	SOF
1	101.04	101.24	100.89	101.60	101.75	100.82	101.45	100.35
2	101.56	96.79	101.53	98.19	100.89	98.99	100.41	99.21
3	100.25	100.16	100.25	100.16	101.04	100.68	99.49	100.56
4	100.62	102.98	100.62	102.56	100.96	101.46	100.54	101.17
5	98.89	98.77	99.01	98.76	99.25	100.21	99.71	100.16
6	101.09	98.39	100.19	98.37	100.45	98.70	100.74	99.54
7	100.04	99.36	100.04	99.37	100.79	99.73	99.91	99.57
8	100.84	100.18	100.81	100.17	99.18	99.07	100.59	98.51
9	101.29	99.35	101.29	98.97	100.38	99.78	100.32	100.57
10	100.89	99.04	100.89	98.98	100.78	99.31	100.48	99.07
11	99.45	99.44	99.58	98.63	99.65	99.40	100.35	100.80
12	100.94	97.11	100.86	98.19	99.56	98.51	101.22	98.61
Mean (%R)	**100.58**	**99.40**	**100.50**	**99.50**	**100.39**	**99.72**	**100.43**	**99.84**
%RSD	**0.776**	**1.728**	**0.712**	**1.414**	**0.810**	**0.924**	**0.557**	**0.893**
RMSEP	**0.0396**	**0.1908**	**0.0384**	**0.1831**	**0.0382**	**0.1438**	**0.0358**	**0.1235**

Table 4. Optimized parameters of ANN.

Method	ANN		GA-ANN	
Drug	LED	SOF	LED	SOF
Architecture	201-10-1	201-7-1	99-4-1	91-4-1
Hidden neurons number	10	7	4	3
Transfer functions	Purelin-Purelin			
Learning rate	0.1	0.1	10	10
Training function	TRAINLM			

agreement with the label claim. The obtained results were statistically compared to those obtained by the reported method [8]. No significant differences were found by applying two tail student *t*-test and F-test at 95% confidence level[16], indicating good accuracy and precision of the proposed methods for the analysis of the studied drug in its pharmaceutical dosage form, as shown in **Table 5.**

CONCLUSION

In this study, accurate and precise multivariate chemometric models were developed. It was found that LED and SOF can be determined simultaneously in their tablets by using the developed methods. The developed methods has the advantages of being sensitive and inexpensive unlike HPLC procedure which is time consuming and expensive.

Table 5. Statistical comparison for the results obtained by the proposed methods and the reported method for the analysis of LED and SOF in Sofolanork plus® tablets.

Method	Drug	Mean	N*	S.D	% RSD	*t***	F**
PLS	LED	100.80	5	0.717	0.718	1.056 (2.306)	3.589 (6.388)
	SOF	99.54		0.685	0.686	1.602 (2.306)	2.682 (6.388)
GA-PLS	LED	100.78		0.664	0.663	1.043 (2.306)	3.071 (6.388)
	SOF	99.65		0.546	0.545	1.204 (2.306)	1.704 (6.388)
ANN	LED	100.41		0.740	0.741	1.753 (2.306)	3.813 (6.388)
	SOF	99.79		0.550	0.549	1.730 (2.306)	1.724 (6.388)
GA-ANN	LED	100.28		0.799	0.797	0.994 (2.306)	4.441 (6.388)
	SOF	99.83		0.537	0.536	1.954 (2.306)	1.644 (6.388)
Reported method [8]	LED	101.61		0.379	0.377	-------	-------
	SOF	101.94		0.419	0.420	-------	-------

* *No. of experimental.*
** *The values in the parenthesis are tabulated values of t and F at (p= 0.05).*
*** *Absorbance subtraction method at which a mathematically estimated factor representing the absorbance ratio (A262.4/A325) for pure LED was calculated, then this factor was used for simultaneous quantitation of LED and SOF using an equation computed at λiso (262.4 nm)[8].*

Application of GA on PLS and ANN models enhance the results with respect to RMSEP. The developed methods can be applied for routine and analysis of ledipasvir in it pure form and in tablets.

Conflict of Interest

The authors declare that they don't have any conflict of interest.

REFERENCES

1. Wikipedia. Available from: https://en.wikipedia.org/wiki/Ledipasvir. Accessed on September 30th, 2016.
2. European Medicines Agency, *Committee for Medicinal Products for Human Use (CHMP) Available from:* http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_Public_assessment_report/human/003850/WC500177996pdf. Accessed on Accessed on September 30th, 2016.
3. Zaman, B.; Siddique, F.; Hassan, W. RP-HPLC method for simultaneous determination of sofosbuvir and ledipasvir in tablet dosage form and its application to in vitro dissolution studies. *Chromatographia* **2016**, 1-9.
4. Elkady, E. F.; Aboelwafa A. A. A rapid and optimized LC-MS/MS method for the simultaneous extraction and determination of sofosbuvir and

ledipasvir in human plasma. *J. AOAC Int*. **2016**; *99* (5), 1252-1259.
5. Chen, Y.; Chen, W.; Zhou, G.; Jin, L.; Zheng, Y.; Lin, W.; Pan, Z. Simultaneous determination of ledipasvir, sofosbuvir and its metabolite in rat plasma by UPLC-MS/MS and its application to a pharmacokinetic study. *J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci.* **2016**, *1008*, 255-259.
6. Bendas, E. R.; Basalious, E. B.; Karim, I. A. Quantification of sofosbuvir and ledipasvir in human plasma by UPLC-MS/MS method: Application to fasting and fed bioequivalence studies. J. *Chromatogr. B: Analyt. Technol. Biomed. Life Sci*. **2016**, *1028*, 63-70.
7. Abo-Talib; Nisreen, F.; El-Ghobashy.; Mohamed R.; Tammam, Marwa H. Spectrophotometric Methods for Simultaneous Determination of Sofosbuvir and Ledipasvir (HARVONI Tablet): Comparative Study with Two Generic Product. *J. AOAC Int.* **2017**. DOI: *https://doi.org/10.5740/jaoacint.16-0330*.
8. Nada, S. A.; Nehal, F. F. Innovative spectrophotometric methods for determination of newly discovered combination for hepatitis C treatment. *Anal. Chem. Lett.* **2016**, *6* (6), 783-794.
9. Brereton, R.G. Multilevel multifactor designs for multivariate calibration, *Analyst* **1997**, (122) 1521-1529.
10. Attia, A. M.; Nassar, M. W.; El-Zeiny, M. B.; Serag, A. Stability indicating methods for the analysis of cefprozil in the presence of its alkaline induced degradation product. *Spectrochim. Acta Part A* **2016**, *159*, 1-6.
11. Attia, A. M.; Nassar, M. W.; El-Zeiny, M. B.; Serag, A. Effect of genetic algorithm as a variable selection method on different chemometric models applied for the analysis of binary mixture of amoxicillin and flucloxacillin: A comparative study. *Spectrochim. Acta Part A* **2016**, *156*, 54-62.
12. Attia, A. M.; Nassar, M. W.; El-Zeiny, M. B.; Serag, A. Firefly algorithm versus genetic algorithm as powerful variable selection tools and their effect on different multivariate calibration models in spectroscopy: A comparative study. *Spectrochim. Acta Part A* **2017**, *170*, 117-123.
13. Haaland, D. M.; Thomas E. V. Partial least-squares methods for spectral analyses. Relation to other quantitative calibration methods and the extraction of qualitative information., *Anal. Chem*. **1988**, *60*, 1193-1202.
14. Abbasi-Tarighat, M.; Shahbazi, E.; Niknam, K., Simultaneous determination of Mn 2+ and Fe3+ as 4, 4′[(4-cholorophenyl) methylene] bis (3-methyl-1-phenyl-1 Hpyrazol-5-ol) complexes in some foods, vegetable and water samples by artificial neural networks. *Food Chem*. **2013**; *138*, 991-997.
15. Afkhamia, A.; Abbasi-Tarighata, M.; Khanmohammadi, H. Simultaneous determination of Co^{2+}, Ni^{2+}, Cu^{2+} and Zn^{2+} ions in foodstuffs and vegetables with a new Schiff base using artificial neural networks. *Talanta* **2009**, *77*, 995-1001.
16. Armitage, P.; Berry, G. *Statistical methods in medical research*. 3rd Edn., Oxford, UK: Blackwell; **1994**.