PREDICTING HYPERGLYCEMIA USING NIR SPECTRUM OF SPENT FLUID IN HEMODIALYSIS PATIENTS**

V. Matović [1*], J. Trbojević-Stanković [1,2], L. Matija [1], D. Sarac [1], A. Vasić-Milovanović [1], A. Petrović [1]

[1] Belgrade University, 11120 Belgrade, Serbia; e-mail: valentinamatovic14@gmail.com
[2] Clinic of Urology, University Hospital Center "Dr Dragiša Mišović – Dedinje", 11000 Belgrade, Serbia

We aimed to assess the near infrared spectroscopy as a method for non-invasive on-line detection of hyperglycemia from spent hemodialysis effluent. We used partial least squares regression and several machine learning algorithms: random forest (RF), logistic regression, K-nearest neighbor (KNN), support vector machine (SVM), decision tree classifier, and Gaussian naive Bayes (NB) to classify normoglycemia from hyperglycemia. These classifier methods were used on the same dataset and evaluated by the area under the curve. The serum glucose levels were presented in the form of a binomial variable, where 0 indicated a glucose level within reference range and 1 a glucose level beyond the normal limit. For this reason, the methods of machine learning were applied as more specific methods of classification. RF and SVM have shown the best classification accuracy in predicting hyperglycemia, while decision tree and NB showed average accuracy.

Keywords: *hemodialysis, machine learning, spent dialysate, near infrared spectroscopy.*

ПРОГНОЗИРОВАНИЕ ГИПЕРГЛИКЕМИИ С ИСПОЛЬЗОВАНИЕМ СПЕКТРА В БЛИЖНЕМ ИК-ДИАПАЗОНЕ ОТРАБОТАННОГО ДИАЛИЗИРУЮЩЕГО РАСТВОРА У ПАЦИЕНТОВ, НАХОДЯЩИХСЯ НА ГЕМОДИАЛИЗЕ

V. Matović [1*], J. Trbojević-Stanković [1,2], L. Matija [1], D. Sarac [1], A. Vasić-Milovanović [1], A. Petrović [1]

УДК 543.42:616-073.27

[1] Белградский университет, 11120 Белград, Сербия; e-mail: valentinamatovic14@gmail.com
[2] Клиника урологии, Университетский центр "Доктор Драгиша Мишович – Дединье", 11000 Белград, Сербия

(Поступила 11 июля 2019)

Спектроскопия в ближней ИК-области использована как метод неинвазивного онлайн выявления гипергликемии из отработанного диализирующего раствора у пациентов, находящихся на гемодиализе. Для отделения гликемии в норме от гипергликемии использованы частичная регрессия наименьших квадратов и алгоритмы машинного обучения: случайный лес (RF), логистическая регрессия, K-ближайший сосед (KNN), метод опорных векторов (SVM), классификатор "дерева решений" и гауссов наивный Байес (NB). Эти методы использованы для одного и того же набора данных.

**Some of the materials from this manuscript were presented at the 56th Congress of the European Renal Association – European Dialysis and Transplantation Association, held in Budapest, Hungary in June 2019.

Full text is published in JAS V. 88, No. 3 (http://springer.com/journal/10812) and in electronic version of ZhPS V. 88, No. 3 (http://www.elibrary.ru/title_about.asp?id=7318; sales@elibrary.ru).

Оценки проведены по площади под кривой. Уровень глюкозы в сыворотке крови представлен в виде биномиальной переменной, где 0 – уровень глюкозы в пределах референтного диапазона, 1 – уровень глюкозы за пределами нормы. Методы машинного обучения применены в качестве более специфических методов классификации. Методы RF и SVM показывают наилучшую точность классификации при прогнозировании гипергликемии, "дерево решений" и NB – среднюю.

Ключевые слова: *гемодиализ, машинное обучение, отработанный диализат, спектроскопия в ближней ИК-области.*

Introduction. Chronic kidney disease (CKD) and diabetes are public health problems that influence millions of people worldwide. Recent epidemiological studies suggest that the global prevalence of diabetes mellitus will grow from 2.8% in the year 2000 to nearly 4.4% in the 2030 [1]. Inadequate blood glucose control is considered the leading cause of diabetic nephropathy and progression to renal insufficiency. Poorly regulated disease, i.e., poor metabolic control, speeds up the process of renal deterioration. Contributing risk factors include duration and type of diabetes, associated hypertension, dyslipidemia and hyperuricemia, and their adequate management can slow the progression of diabetic nephropathy [2]. Long-term exposure to hyperglycemic milieu is the primary cause of most diabetic complications, including diabetic nephropathy [3, 4].

The most-studied biological fluids of clinical interest are blood, urine, and, recently, spent dialysate. Eddy and Arnold have shown the possibility of glucose detection in spent dialysate utilizing near infrared spectroscopy (NIRS) [5]. Due to the large amount of solution required to perform a single dialysis procedure, the dialysis fluid is created by on-line mixing of the ultrapure water with an electrolyte concentrate inside a dialysis machine. The machine guarantees electrolytic composition, pH, temperature, and flow rate of the dialysis fluid. This fluid usually contains a high glucose concentration to minimize the nutritional loss that occurs in patients during hemodialysis [6, 7].

The direct way of determining metabolites in the patient's blood during dialysis is based on a simple multicomponent determination of certain substances such as urea, creatinine, glucose, electrolytes, etc. However, this process requires taking blood from already anemic patients. A possible safer and more convenient alternative would be on-line monitoring of substances using biosensors. On-line monitoring of urea, creatinine, or glucose is complicated by the fact that blood is a highly saturated fluid, prone to clotting. Kaiser et al. have highlighted the problems that arise in the measurement of blood glucose as a result of the blood matrix complexity [8]. All amperometric biosensors suffer from interferences in complex matrices such as blood or serum. Glucose biosensors also suffer from rapid performance deterioration after implantation due to tainted surfaces and coagulation caused by poor biocompatibility. Coating of the sensor with protein or cellular material from the biological matrix is also a frequent phenomenon [9]. We therefore hypothesized that spectrum analysis of the spent dialysate is more reliable as a hyperglycemia sensor.

Experimental. Samples of spent dialysate were obtained upon five consecutive dialysis treatments from three patients treated with chronic bicarbonate hemodialysis (HD). Inclusion criteria were a stable hemodialysis prescription, stable intradialytic blood pressure, absence of physical weakness or dyspnea, and ability to rest in a 45–90° position during the entire dialysis session. The patients were well compliant with dialysis procedures and had no active infections nor intradialytic complications at the time of sampling. All HD treatments were performed under the usual protocol, including ultrafiltration rates prescribed to remove the interdialytic weight gain. Patients enrolled in the study were routinely treated with dialyzers containing the PAES high-flux membrane (Polyflux 170H, Gambro) with a membrane surface area of 1.7 m^2. Dialysis was performed using a Dialog+Adimea (BBraunAvitum AG, 34209 Melsungen, Germany) machine. The dialysate contained (mmol/L) Na^+ 138, Cl^- 110.5, K^+ 2, Ca^{++} 1.75 or 1.50, Mg^{++} 1, CH_3COO^- 3, and glucose 1 g/L. The mean dialysate flow was 500 mL/min, and mean effective blood flow was 300 mL/min. All patients were dialyzed via antebrachial arteriovenous fistulas using a two-needle system. The Ethical committee of the University Hospital Center Dr Dragiša Mišović reviewed the study protocols and all patients provided written informed consent before participating in the study.

Samples of spent dialysate were collected directly from the dialyzer outlet during the dialysis session. It was previously ensured that the dialysate flow was free and uninterrupted. The spent dialysate, containing dialyzed waste metabolites, flowed upwards through the cartridge and the outlet to the external environment. The baseline dialysate samples were taken from the effluent line 15-min after the beginning of the dialysis session. At the same time, blood samples were taken from dialysis blood line, blood pressure was measured, and patient's position was registered. Blood samples were collected from the arterial line of the dialysis system, e.g., coming from the patient immediately before entering the dialysis circuit. Blood sampling was nec-

essary because we had to know the numerical value of glucose in the patient's blood for the training of the machine learning algorithm. For each sample, 15 mL of spent dialysate solution was collected into a container and stored at room temperature for approximately 3 h before being transported to the research laboratory.

NIR absorbance spectra of the samples were measured the day after HD treatment. The absorption spectrum of each sample was measured three times. This provided a dataset with 126 spectra. NIR optical absorption spectra were registered using a spectrometer Lambda 950 (Perkin Elmer). This spectrometer was equipped with a standard tungsten halogen lamp and PbS detector. The wavelength region of interest was 700-1700 nm and resolution set to 2 nm. The instrument was connected to a PC running the Windows 7 operating system and controlled by a Perkin Elmer UV WIN LAB Explorer. Serum glucose was measured using the Dimension RxLMax (Siemens Healthcare GmbH, Germany) machine. The assay is based on the hexokinase method. Glucose level above 6 mmol/L was considered hyperglycemic [10].

For data analysis, we used partial least squares regression (PLSR) and several machine learning (ML) algorithms: random forest (RF), logistic regression, K-nearest neighbor (KNN), support vector machine (SVM), decision tree classifier, and Gaussian naive Bayes (NB). These classifier methods were used on the same dataset with cross validation and the area under the curve (AUC) evaluation. Results were compared between different methods. Spectra were normalized using standard normal variates (SNV). In constructing the model, we chose the classification approach. For all patients, the serum glucose level was divided into two groups: normal (0) and hyperglycemic (1).

The decision tree represents a classifier that separates variables and their numeric values in such a way that there is the greatest difference between values in features relative to the target variable.

RF utilizes an ensemble of decision trees in order to reach a conclusion. Additionally, it applies random samples and features from the dataset for every tree. Every decision tree provides solution, and results from a large number of trees should converge to the most accurate results.

KNN classifies data based on the nearest point in multidimensional space, which are established in the training phase. Best results were acquired when four (K = 4) neighbors were used. The distance between points was calculated in Euclidian metric.

SVM utilizes data points to the support plane, which separates data according to the target variable. The separation plane is supported with data points with the greatest margin.

Logistic regression is a classifier that uses a sigmoid function in order to separate the binomial target variable based on feature values. C was set to 0.1. Solver was set to 'lib linear.'

The NB classifier is based on the Bayes theorem, and prior probability was set to 'None.'

Stratified cross-validation (CV) with 10 folds was used. The term 'Stratified' suggests that each division of data on the train and test set was created in such a way that class proportions (in this case 0 and 1) are roughly the same size in each fold. Algorithms were evaluated using the obtained operating characteristic (ROC) curve and AUC score. The ROC curve coupled with its AUC is a common method used to estimate the diagnosis potential of a classifier in clinical applications. A larger AUC indicates higher prediction ability.

Results and discussion. The correlation coefficient using PLSR technique was observed to be 0.503. The nonlinearity between data and blood glucose was due to the presence of several components with overlapping spectral features. Furthermore, the value of R^2 for PLSR technique was very low; therefore ML was considered in the present work for estimation of blood glucose. According to the ROC curve and AUC criterion, the best model was acquired using the RF algorithm, followed by SVM and LR. RF gave an accuracy of 91%, followed by SVM of 89% and LR of 82%. KNN had an accuracy of 80%, while Decision Tree with AUC and NB had poor accuracy of 71% and 58% respectively. Results are presented in Fig. 1. The AUC score of all classifiers are presented on Fig. 2.

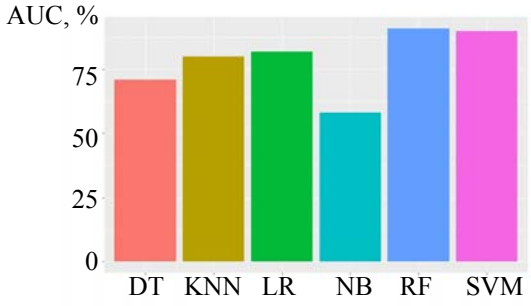

Fig. 1. AUC score of algorithms.

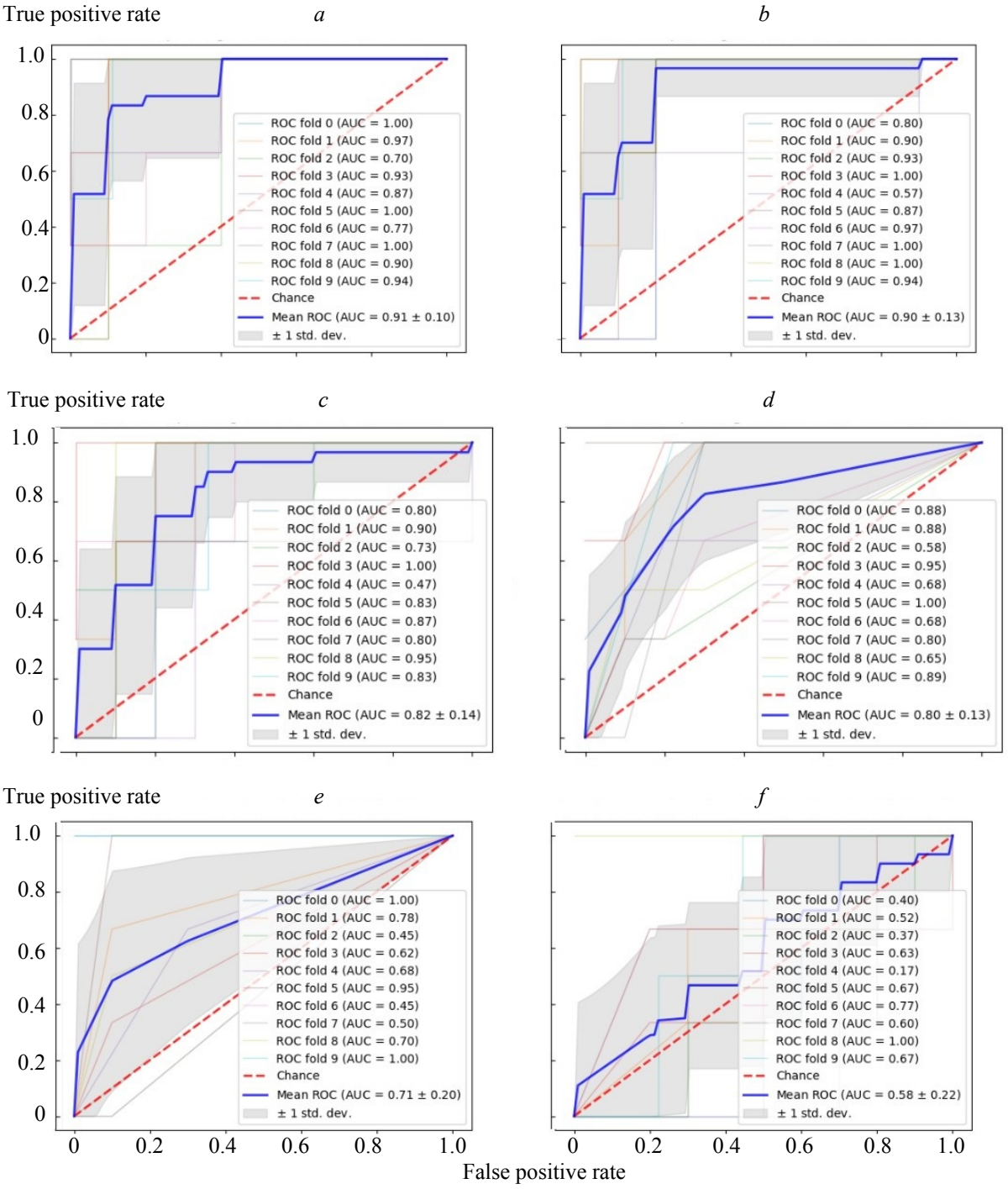

Fig. 2. AUC score of RF (a), SVM (b), LR (c), KNN (d), DT (e), and Gaussian NB (f) algorithms.

Diabetes complications can be avoided by frequent blood glucose monitoring. For patients using insulin, pre- and post-dialysis blood glucose self-monitoring is recommended at each dialysis session [11]. In these patients, due to chronic anemia and such excessive sampling, we tend to reduce the blood taken to the lowest possible measure.

In HD patients with poorly controlled diabetes, the blood glucose levels differ widely between non-dialysis and dialysis days [12]. Hyperglycemia on hemodialysis is caused by the dialyzer adsorption of previously administered insulin [12]. For patients with poor glycemic control and high blood glucose levels before starting hemodialysis treatment, blood glucose levels fluctuate widely.

In this study we assessed blood glucose level using the spectrum of the spent dialysate. Several ML algorithms were then applied in order to classify samples in the normal and hyperglycemic groups. The applied PLSR regression technique yielded poor results, which was expected, since the basic assumption of these methods is the independence of various parameters, which is not the case for a non-Newtonian fluid such as spent dialysis fluid. The goal of this study was to create a classifier that would be able to classify patients with normal and high blood glucose levels based on the spent dialysis fluid spectrum. Then the prediction becomes a binary (yes/no) classification problem. From a computational standpoint, classification problems are easier and more efficient to solve than regression. Data obtained from NIR spectrum of the spent dialysate can be input into predictive models to predict hyperglycemic events. Glucose measurement techniques need to allow for constant monitoring of glucose, need to be non-invasive, and need to focus on detection of blood glucose concentration when reaching hyperglycemic levels. Non-invasive methods for monitoring glucose level based on infrared spectroscopy were first invented during the nineties [13]. Since then, a wide range of techniques has been developed for the non-invasive observation of glucose based on chemical, optical, and electrical techniques using microsensor and computer technologies [14]. This development of non-invasive techniques was preceded by successful *in vitro* studies that were based on the determination of glucose in aqueous solutions [15, 16] or whole blood [17] by NIRS. Studies were mainly based on the effects of glucose on certain secondary processes. One of the most famous examples is the effect of glucose on the scattering properties of tissue. However, propagation of light through tissue is complicated by the heterogeneous nature of the tissue matrix, thus creating a problem [18]. The NIR region of the electromagnetic spectrum covers the wavelength range between 750 and –2500 nm [19]. Light absorption from the NIR region is primarily caused by the presence of functional groups, C-H, O-H, and N-H. They absorb protons of a certain frequency belonging to the NIR region. The principle that the absorption pattern of NIR light (700–1700 nm) can be quantitatively related to the glucose concentration is proven [20–25].

Among all available methods, PLSR regression has been used most widely for the analysis of the NIR spectral data [26]. The biggest problem with PLSR methods is that the spectrum property relationship is assumed to be linear. However, this premise cannot be applied to systems with strong intermolecular or intramolecular interactions. If we measure the amount of glucose in a fluid that contains other substituents, we cannot simply apply the Beer–Lambert law because we have interactions between components, incorrect distribution of fluid components, and baseline shift. All of this leads to nonlinearity of the system. This makes nonlinear calibration methods necessary for building robust calibration models since these methods have the potential to model heavy intrinsic nonlinearities that are found in natural multicomponent systems.

ML has also been applied to non-invasive glucose measurements (NIGM) in various ways. The researchers combined ML to investigate glucose level in patient blood [24, 27]. ML methods have not only been applied in the tracking of glucose but also in predicting hypoglycemia [28, 29]. RF, SVM, KNN, and NB were used by Sudharsan et al. [30] to predict hypoglycemia, whereas support vector regression was used by Georga et al. [31] for the same reason. Similarly, in the case of neuropathy, Du Brava et al. used RF in order to select specific features targeting prediction of diabetic peripheral neuropathy (DPN) [32].

The research conducted by Roth et al. [33] has already shown a correlation between the NIR spectrum and the glucose concentration of spent dialysis fluid. On the other hand, the goal of the study presented in this paper has been to demonstrate a correlation of the NIR spectrum of spent dialysis fluid and blood glucose concentration. Further steps in this ongoing study will involve the determination of glucose concentration in spent dialysis fluid and blood and their interrelation with the NIR spectrum.

In this study, we did not record the occurrence of hypoglycemia in our patients, although its detection is clinically more important than the detection of hyperglycemia. Obviously, it would not be ethical to deliberately induce the state of hypoglycemia to test the success of our algorithm in this regard. On the other hand, detecting incidental hypoglycemia would require unpredictable amounts of time and resources. Still, we believe that the presented algorithm would be equally successful in detecting hypoglycemia as it was in identifying hyperglycemia.

Conclusions. The aim of this study was to detect hyperglycemia from the matrix of the spent dialysate fluid using NIRS. Additionally, further details of the suitability of particular data mining methods used to detect hyperglycemia, such as RF, logistic regression, KNN, SVM, NB, and decision tree classifier, have been described. RF and SVM have shown the best classification accuracy for the prediction of hyperglycemia, while decision tree and NB have shown average accuracy. This approach can be used for on-line prediction of hyperglycemia in hemodialysis patients. Further studies should assess the reliability of NIR with the ML

technique for classification problems of a whole range of molecules that can be detected in the spent dialysis fluid.

Acknowledgments. This study was supported by the Ministry of Science and Education of Republic of Serbia (Project III45009 and III41006).

REFERENCES

1. S. Wild, G. Roglic, A. Green, R. Sicree, H. King, *Diabetes Care*, **27**, 1047–1053 (2004).
2. T. H. Hostetter, *N. Engl. J. Med*., **351**, No. 13, 1344–1346 (2004).
3. D. M. Nathan, S. Genuth, J. Lachin, P. Cleary, O. Crofford, M. Davis, L. Rand, C. Siebert, *N. Engl. J. Med*., **329**, 977–986 (1993).
4. J. Pirart, *Diabetes Care,* **1**, 168–188 (1978).
5. C. V. Eddy, M. A. Arnold, *Clin. Chem*., **47**, 1279–1286 (2001).
6. F. M. Parsons, W. K. Stewart, In: *Replacement of Renal Function by Dialysis,* Springer, 148–170 (1983).
7. R. Sam, M. Vaseemuddin, W. H. Leong, B. E. Rogers, C. M. Kjellstrand, T. S. Ing, *Hemodial. Int*., **10**, 15–28 (2006), doi:10.1111/j.1542-4758.2006.01170.x.
8. N. Kaiser, *Method for Determining the Contents of Metabolic Products in the Blood,* U. S. Patent 4, 169, 676, iss. October 2 (1979).
9. T. Koschinsky, L. Heinemann, *Diabetes. Metab. Res. Rev*., **17**, 113–123 (2001).
10. *American Diabetes Association Clinical Practice Recommendations*, American Diabetes Association (1997).
11. T. Nakao, M. Inaba, M. Abe, K. Kaizu, K. Shima, T. Babazono, T. Tomo, H. Hirakata, T. Akizawa, *Ther. Apher. Dial*., **19**, 40–66 (2015).
12. M. Abe, K. Kalantar-Zadeh, *Nat. Rev. Nephrol*., **11**, 302 (2015).
13. M. A. Arnold, *Curr. Opin. Biotechnol*., **7**, 46–49 (1996).
14. J. García-Guzmán, N. González-Viveros, H. H. Cerecedo-Núnez, In: *Emerging Challenges for Experimental Mechanics in Energy and Environmental Appl. Proc. 5th Int. Symp. Optics in Industry (ISEM-SOI)*, 2015, Springer, 55–63 (2017).
15. K. E. Kramer, G. W. Small, *Vib. Spectrosc*., **43**, 440–446 (2007).
16. A. K. Amerov, J. Chen, G. W. Small, M. A. Arnold, *Anal. Chem*., **77**, 4587–4594 (2005).
17. Q.-B. Li, L.-N. Li, G.-J. Zhang, *Infrared Phys. Technol*., **53**, 410–417 (2010).
18. M. A. Arnold, G. W. Small, *Anal. Chem*., **77**, 5429–5439 (2005).
19. S. F. Malin, T. L. Ruchti, T. B. Blank, S. N. Thennadil, S. L. Monfre, *Clin. Chem*., **45**, 1651–1658 (1999).
20. K. Maruo, M. Tsurugi, J. Chin, T. Ota, H. Arimoto, Y. Yamada, M. Tamura, M. Ishii, Y. Ozaki, *IEEE J. Sel. Top. Quantum Electron*., **9**, 322–330 (2003).
21. C. Araujo-Andrade, F. Ruiz, J. R. Martínez-Mendoza, H. Terrones, *AIP Conf. Proc*., 234–239 (2004).
22. R. Liu, W. Chen, Y. Chen, K. Xu, *Opt. Diagnostics Sens. VIII, Int. Soc. Opt. Photon*., 68630Q (2008).
23. K. Xu, Q. Qiu, J. Jiang, X. Yang, *Opt. Lasers Eng*., **43**, 1096–1106 (2005).
24. C. S. Soh, X. Zhang, J. Chen, P. Raveendran, P. H. Soh, J. H. Yeo, A*dv. Biomed. Clin. Diagnostic Syst. VI, Int. Soc. Opt. Photon*., 68480B (2008).
25. E. Guevara, F. J. González, *Rev. Mex. Fís.*, **56**, 430–434 (2010).
26. D. M. Haaland, E. V. Thomas, *Anal. Chem*., **60**, 1193–1220 (1988).
27. P. Zuo, Y. Li, J. Ma, S. Ma, *ICNN&B'05. Int. Conf., IEEE*, 1350–1353 (2005).
28. K. Y. Chan, S.-H. Ling, T. S. Dillon, H. T. Nguyen, *Expert Syst. Appl*., **38**, 9799–9808 (2011).
29. S. Malik, R. Khadgawat, S. Anand, S. Gupta, *Springerplus,* **5**, No. 1, 1–12 (2016).
30. B. Sudharsan, M. Peeples, M. Shomali, *J. Diabetes Sci. Technol*., **9**, 86–90 (2014).
31. E. I. Georga, V. C. Protopappas, D. Ardigò, D. Polyzos, D. I. Fotiadis, *Diabetes Technol. Ther*., **15**, 634–643 (2013).
32. S. DuBrava, J. Mardekian, A. Sadosky, E. J. Bienen, B. Parsons, M. Hopps, J. Markman, *Pain Med*., **18**, 107–115 (2017).
33. A. Roth, F. Dornuf, O. Klein, D. Schneditz, H. Hafner-Gießauf, W. Mäntele, *Anal. Bioanal. Chem*., **403**, 391–399 (2012).