chemosensors

MDPI

Article

SERS Sensor for Human Glycated Albumin Direct Assay Based on Machine Learning Methods

Ekaterina A. Slipchenko [1], Irina A. Boginskaya [1,*], Robert R. Safiullin [2], Ilya A. Ryzhikov [1,3], Marina V. Sedova [1], Konstantin N. Afanasev [1], Natalia L. Nechaeva [4], Ilya N. Kurochkin [4], Alexander M. Merzlikin [1] and Andrey N. Lagarkov [1]

[1] Institute for Theoretical and Applied Electromagnetics RAS, Moscow 125412, Russia
[2] Moscow Institute of Physics and Technology, Dolgoprudny, Moscow 141700, Russia
[3] FMN Laboratory, Bauman Moscow State Technical University, Moscow 105005, Russia
[4] Emanuel Institute of Biochemical Physics RAS, Moscow 119334, Russia
* Correspondence: i.boginskaya@bk.ru

Abstract: In this study, a non-labeled sensor system for direct determining human glycated albumin levels for medical application is proposed. Using machine learning methods applied to surface-enhanced Raman scattering (SERS) spectra of human glycated albumin and serum human albumin enabled the avoidance of complex sample preparation. By implementing linear discriminant analysis and regularized linear regression, classification and regression problems were solved based on the spectra obtained as a result of the experiment. The results show that, coupled with data augmentation and a special cross-validation procedure, the methods we employed yield better results in the corresponding tasks in comparison with popular random forest methods and the support vector method. The results show that SERS, in combination with machine learning methods, can be a powerful and effective tool for the simple and direct assay of protein mixtures.

Keywords: human serum albumin; surface-enhanced Raman spectroscopy; protein mixtures; machine learning methods; non-labeled sensor system; medical sensor system; glycated human serum albumin assay

Citation: Slipchenko, E.A.; Boginskaya, I.A.; Safiullin, R.R.; Ryzhikov, I.A.; Sedova, M.V.; Afanasev, K.N.; Nechaeva, N.L.; Kurochkin, I.N.; Merzlikin, A.M.; Lagarkov, A.N. SERS Sensor for Human Glycated Albumin Direct Assay Based on Machine Learning Methods. *Chemosensors* **2022**, *10*, 520. https://doi.org/10.3390/chemosensors10120520

Academic Editor: Lucio Litti

Received: 11 November 2022
Accepted: 4 December 2022
Published: 7 December 2022

1. Introduction

Surface-enhanced Raman scattering (SERS) is an effective method for the quantitative and qualitative determination of complex biological objects, in particular proteins [1,2].

To carry out research using the SERS method, it is necessary to use special substrates that realize the effect due to the surface features on which the phenomenon of plasmon resonance and the localization of the electromagnetic field occur. Such substrates are most often colloidal solutions of nanoparticles [3–7] or various planar substrates based on noble metals [2,8] and lithographic structures [1,9–12]. In particular, planar extended nanostructured substrates based on silver and gold, obtained by vacuum evaporation methods, were often effective for detecting proteins [2,13,14]. They have proven to be reliable, reproducible, optically stable structures for implementing various optical effects, which is important for the prospect of using such structures for optical and plasmonic applications [15–18].

One urgent task is the determination of glycated human serum albumin (GHSA) in blood plasma, as the level of albumin glycation in the human body is an integral feature characterizing average sugar content over time intervals of 2–3 weeks [19] as a marker for type 2 diabetes [20].

There are a number of methods for GHSA detection, including ion-exchange high-performance liquid chromatography (HPLC), boronate affinity chromatography, immunoassays (radioimmunoassay and enzyme-linked immunosorbent assay), a colorimetric method with thiobarbituric acid, and enzymatic methods using proteinase and ketamine oxidase [21–24]. All of these methods require labor-intensive sample preparation, the use of many reagents,

and complex equipment, and they cannot be used in in-patient facilities. The use of SERS or Raman spectroscopy in combination with mathematical methods can facilitate the detection of GHSA in both qualitative and quantitative determination of complex biological analytes due to the simplification of sample preparation and the automation of measurements and result processing [25,26]. In one study, a partial least squares regression (PLS) calibration model was developed for direct measurement of the Raman spectra of GHSA with a given concentration in a dried precipitate [25]. Another study used SERS spectra of a modified SERS substrate by 4-mercaptophenylboronic acid to develop a calibration curve based on PLS due to covalent binding with GHSA molecules from HSA–GHSA mixtures [26].

The choice of mathematical processing approach determines the possibility of solving the above-described problem due to restrictions in the methods that can be used to analyze spectral data. In particular, it is not always possible to determine spectra groups in new spaces using PCA, as this method cannot fully describe data only by their covariance. In this case, supervised methods, such as linear discriminant analysis (LDA), are more advantageous because they not only take into account data dependencies but also ensure that spectra intensities align with relevant concentrations [27]. PLS and the random forest model (RF) [28], as well as the support vector machine (SVM) method [29], are often used to construct calibration; however, these approaches can also be applied to classification. Although these methods have been demonstrated to be moderately successful in solving these problems, they have a large variance in heterogeneous data. LDA and ridge regression can be employed to achieve stability and reduce variation, respectively.

In this work, we explore the possibility of non-labeling sensor system development applying linear models on processed SERS spectra of HSA–GHSA mixtures in biologically significant concentrations for the classification and quantification of GHSA via calibration curve at direct drying without additional sample preparation, which is an important factor for simplifying the analysis process. For SERS realization, we use thin film silver substrates formed by vacuum sputtering technology and characterized by nanostructured self-organizing controlled morphology [2]. We also show the advantages of the linear model in comparison with indicative models as RF и SVM in terms of their performance on our experimental data (SERS).

2. Materials and Methods

2.1. Sample Preparation

Freshly prepared water solutions of human serum albumin (HSA, Sigma Aldrich, Burlington, MA, USA, LOT#SLBK6136V) and glycated human serum albumin (GHSA, Sigma Aldrich, Burlington, MA, USA, LOT #SLBT1722) were used with a total concentration of 1 g/L of proteins in deionized water from Milli-Q system (Merck KGaA, Darmstadt, Germany). According to [21], the average number of glycation sites per molecule is 1.97. Protein mixtures with a mass content of GHSA in solutions of 0%, 3%, 5%, 7%, 8%, 10%, 11%, 13%, 15%, 18%, 20%, 23%, 25%, and 100% GHSA were prepared from solutions of HSA and GHSA. The total protein concentration remained constant and equal to 1 g L-1. According to typical physiological values of GHSA, the concentration range selected for analysis should be less than 25% [25] of the total albumin value. Concentrations of GHSA 0%, 3%, 5%, 7%, 10%, 13%, 15%, 18%, 20%, 23%, and 25% were used for calibration curve plotting, and concentrations of GHSA 8% and 11% were used for validation.

2.2. SERS Substrate Preparation

Silver thin films were formed according to [2]. To obtain SERS substrates, silver was evaporated and applied to microscope glass slides (Heinz Herenz GmbH, Ulm, Germany) using the 8 kW e-beam evaporator (Quartz Ltd., Kaliningrad, Russia) with a base pressure lower than 5×10^{-6} Torr. The glass slides were preliminarily rinsed with isopropanol (99.5%; Sigma-Aldrich, Burlington, MA, USA) and pre-cleaned in plasma in a residual atmosphere at a pressure of 10^{-3} Torr in the vacuum chamber. All of the films were grown using high-purity 4N (≥99.99%) granulated silver (OOO «Moscow special alloys processing plant», Russia) with

a grain size of 3 mm. The chamber pressure was 5×10^{-6} Torr, the cathode voltage was 12 V, the beam current was 30 mA, and the source accelerating voltage was 8 kV. The thickness of the films was optically controlled during the deposition process. After deposition, the film thickness was selectively measured using a scanning interferometer New View 7200 (Zygo, Middlefield, CT, USA). The film thickness was estimated at 120 nm.

2.3. AFM Analysis

AFM measurements were performed using a Solver Pro (NT-MDT, Zelenograd, Russia) microscope in tapping mode. Images were processed in Gwyddion 2.60 (CMI, Brno, Czech Republic). Morphology parameter statistical analysis was performed using Gwyddion built-in program modules, calculating roughness parameters based on ISO 21920-2:202.

2.4. Protein Solution Deposition

The aliquots of HSA and GHSA solution and HSA–GHSA mixtures were deposited in a volume of 3 µL on the SERS substrate surface. The drops were air-dried without additional action, and the SERS spectra were measured in the formed dry residue, which resembles a coffee ring, as shown in Figure 1.

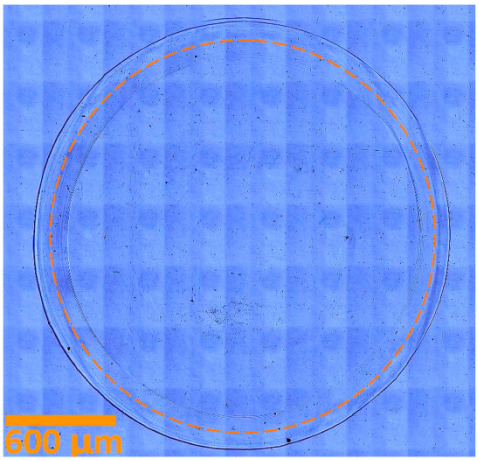

Figure 1. Coffee-ring HSA solution 1 g L-1. All other solutions have a similar appearance. The dotted line marks the area where the spectra were measured.

2.5. Spectra Measurement

SERS spectra measurements were conducted using a confocal Raman spectrometer Alpha 300 R (WITec, Ulm, Germany) based on the confocal microscope with Epiplan Neofluar 50X/0.8 DIC ∞/0. The laser wavelength was 785 nm, the power was 45 mW, and the accumulation time was 15 sec. More than 30 spectra were measured for each sample. The integrational time and total count of spectra were determined by the signal-to-noise ratio (SNR). It can be seen in Figure 2 that this ratio ceases to change noticeably when the number of measured spectra is greater than 15.

From the above, it can be concluded that a further increase in the number of measured spectra does not lead to a change in SNR. Figure 2 shows that the number of spectra, after which the graph reaches a plateau, is low. At the same time, the spectrum measurement time is only 15 sec, which makes the proposed method promising in terms of implementation time, which is important for medical applications.

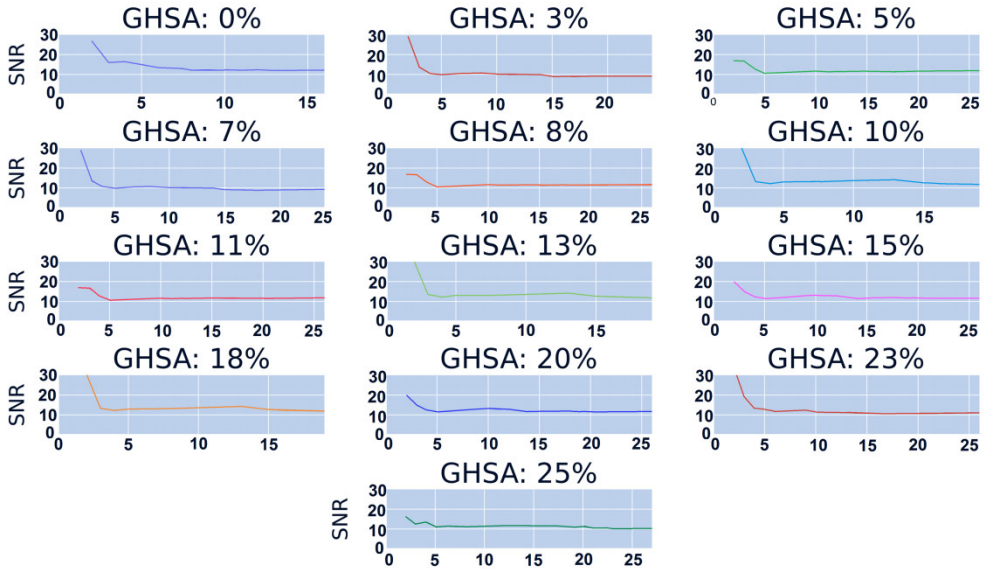

Figure 2. Signal-to-noise ratio for smoothed spectra vs. their quantity by GHSA concentration.

2.6. SERS Spectra Preprocessing

When designing models based on machine learning, data preprocessing is one of the most important stages. Depending on the selected actions, the final quality of the model is determined based on experimental data. The following steps were used to preprocess the dataset containing a total of 393 spectra. First, the Raman shift range (from 400 to 1800 cm^{-1}) was selected. Then, the baseline was corrected using a rubber-band correction. Existing outliers in 70 spectra, such as from abnormal intensity range or ranges strongly out of the general spectra distribution, were manually eliminated. After that, each spectrum was normalized to its own mean and standard deviation. Smoothing was performed using a Savitsky–Golay filter with a window size of 15 and a polynomial order of 3. The processing of the obtained spectra and the subsequent analysis, as well as the construction and optimization of the model parameters, were performed using Python 3.

2.7. Training and Testing Data

The resulting dataset consists of thirteen different mixtures, which were evenly distributed (including concentrations deferred for validation). Samples of eleven mixtures were used to build the model. Among them, 80% were chosen to train the model, and the remaining 20% were used to evaluate its efficiency. Both training and test sets were stratified according to the concentrations. The training dataset was augmented by adding normal noise [30]. Samples from the two remaining mixtures (8% and 11%) were used for validation.

2.8. Machine Learning Algorithms

Two datasets consisted of an X (objects-features) and Y (class labels for classification and GHSA concentration for regression) matrix. Predictive models, namely LDA [11] and linear regression with L1 regularization for feature selection and L2 regularization for weight regulation (LR) [31] were chosen to solve a classification problem and construct the robust regression curve. The regularization parameters (L1 and L2) were selected using cross-validation via grid search, wherein the test set entirely consisted of one type of object (one concentration of GHSA for the model).

2.9. Model Evaluation

The models were compared with the popularly used SVM and RF [29], with parameters also selected on cross validation. A polynomial kernel function with a parameter (gamma) equal to 0.5 was chosen for the support vector method. L2 regularization was also used for this method. The random forest model consisted of 700 trees with unrestricted tree depth,

as well as a minimum number of objects in a leaf equal to 2. This choice allows us to obtain an ensemble of complex basic algorithms with a small offset. The high variance is leveled by the number of trees. To quantify the quality of the results, we used precision with recall metrics for classification and the coefficient of determination with the mean squared error of prediction (RMSE) for regression. They are briefly described below.

2.9.1. Coefficient of Determination R^2

The coefficient of determination, R^2, is used to analyze how differences in one variable can be explained by a difference in a second variable. It could be interpreted as a percent of how many data points fall within the results of the line formed by the regression equation.

2.9.2. Root Mean Square Error (RMSE)

The standard quadratic error function and, at the same time, a quality metric is commonly used in data analysis. Due to the deviation square, this metric is sensitive to outliers, which additionally helps to validate the result.

2.9.3. Precision

Precision shows how many elements selected with the help of a classification model are relevant, i.e., how many actually predicted label concentrations of spectra really correspond to them.

2.9.4. Recall

Recall shows the proportion of correctly selected (by the model) concentrations among all samples with such a concentration value from the experiment

2.9.5. F1 Score

This metric is the harmonic mean of precision and recall, equally taking into account the values of both.

The contribution of sample augmentation to the final result was also evaluated.

3. Results

3.1. SERS Substrate

Figure 3 illustrates an AFM image of the used substrate.

Previous works have described substrates of the same type [2,13]. It should be noted that the general morphology shown in Figure 3a,c demonstrates the typical polycrystalline microstructure of metal films formed by electron beam evaporation on a substrate with a different crystal lattice [18]. The cross-section of the surface is shown in Figure 3b,d. It is clear that the surface contains nanoscale inhomogeneity. Morphology statistical parameters were calculated based on 25 surface cross-sections: root mean square roughness (R_q = 3.6 nm); average third-highest peak to third-lowest valley height (R_{3z} = 12.1 nm); kurtosis (R_{ku} = 4.1); and average wavelength of the profile (λ_a = 198.4 nm). The R_q and λ_a parameters show that whereas the film is very smooth on macroscale, on the sub-micro scale, it is characterized by local inhomogeneity, represented by a change in height and expressed by the parameters R_{3z} and R_{ku}. According to the SERS theory, the mechanism of the effect is determined by the optical parameters of the substrate material and surface roughness [32]. The localization of the electromagnetic field can occur on nanostructured inhomogeneities of the surface [26]. The electromagnetic mechanism typically dominates in SERS, arising from the interaction of the optically excited collective electron oscillation and the analyte molecule. Coupling the laser beam and conductive electrons considerably affects local electromagnetic field distribution in the proximity of the nanostructured metallic surface, increasing Raman scattering cross-section and thereby improving the output signal. From the cross-sectional drawings of the surface in Figure 3b,d, we can see that our substrate is characterized by a complex relief with numerous valleys and heights that can work to localize the electromagnetic field. On the whole, the substrate can realize

the multi-component effect of chemical amplification and electromagnetic amplification, which cannot be separated from each other in this case. The existence of chemical SERS is due to the nature of the silver coating, which interacts with sulfur-containing amino acids, negatively charged amino acids, and amino acids containing an aromatic group [13]. This should facilitate the transfer of charge from the substrate to the analyte molecule.

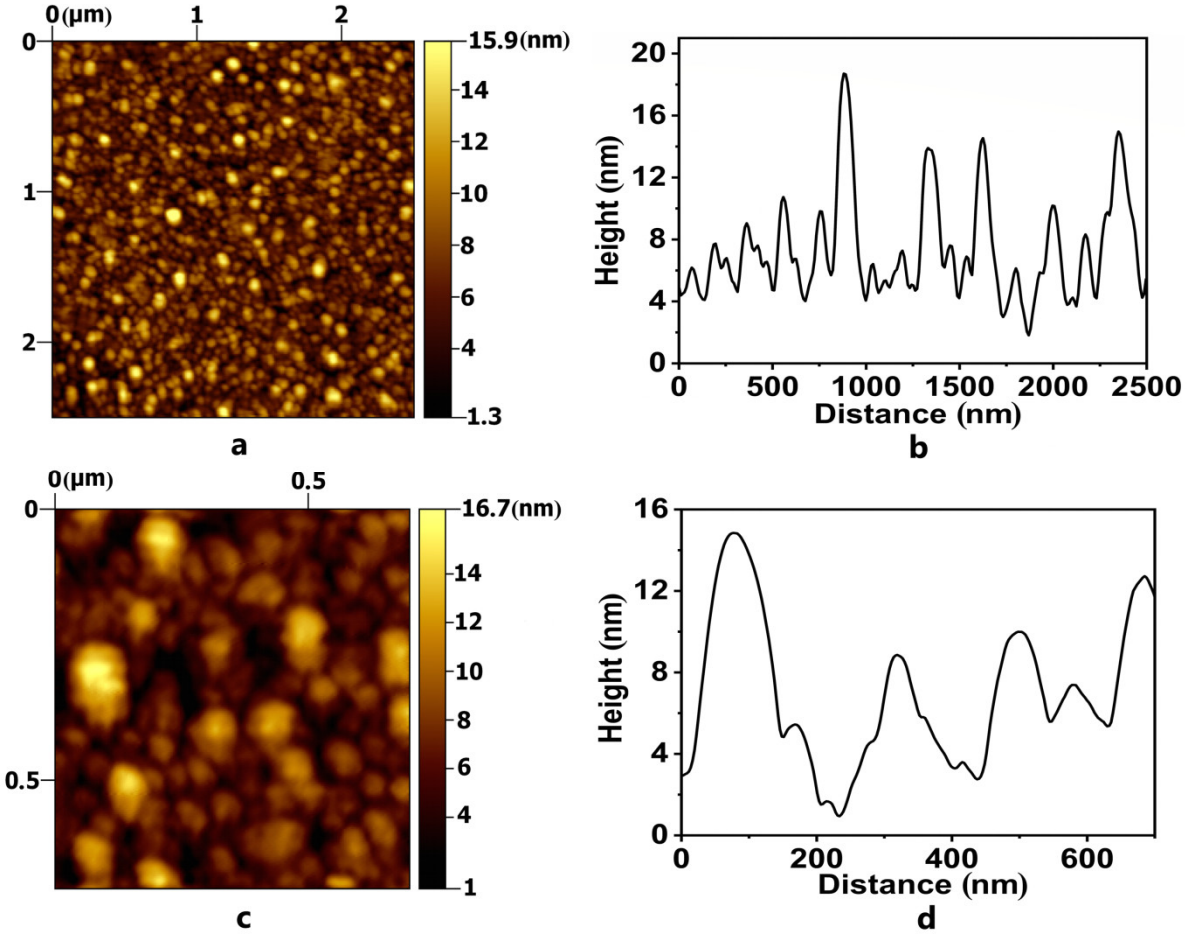

Figure 3. AFM results: (**a**) AFM image of 2.5 × 2.5 μm SERS substrate surface; (**b**) cross-section of 2.5 × 2.5 μm surface; (**c**) AFM image of 0.7 × 0.7 μm SERS substrate surface; (**d**) cross-section of 0.7 × 0.7 μm surface.

Previous experimental and theoretical studies of the SERS properties of semi-continuous metal films characterized as active SERS substrates [33–36] using near-field, atomic force, and electron microscopy, together with a powerful theoretical apparatus [37], made it possible to theoretically and experimentally visualize the localization of an enhanced electromagnetic field on substrate surfaces. In the works mentioned above, the localization of the field appears to be inhomogeneously dispersed over the entire surface of the substrate and, at the same time, localized in the inhomogeneities of the surface morphology. However, the size of such regions is much smaller than the exciting laser beam of the optical system of the microscope. We can assume that the measured spectra represent, on the whole, the average value of the gain from a set of many hotspots that fall into the microscope lens at once. Under the conditions of our measurements, a fairly large region larger than 1 μm falls into the measured beam during the measurement of each spectrum.

An analysis of the surface morphology of the substrates used by us in the present work suggests a similar behavior of the field localization as described. The admissibility of this assumption is based on the fact that the substrates created by us can be represented as a combination of two layers of silver: a semi-continuous silver film on a continuous silver

film. The existence of the first is due to the presence of roughness, and the second is bulk silver, which provides sufficient heat removal to ensure that the substrate remains intact under the conditions of our measurements. This allows us to use higher powers, which in turn leads to a reduction in measurement time, thus providing a high-quality spectrum for a short accumulation time.

Thus, we can state that a complex interaction of the analyte with the substrate can be observed on our substrates, which leads to a noticeable enhancement of the spectra.

3.2. SERS Spectra

Figure 4 shows the obtained spectra of 100% pure HSA and GHSA.

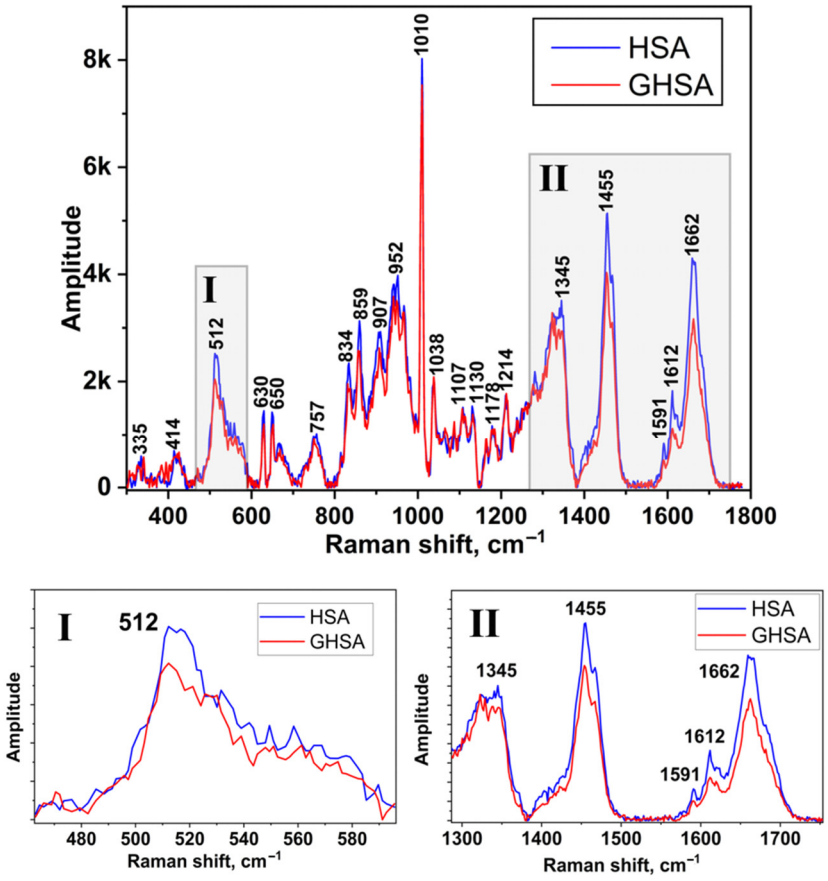

Figure 4. GHSA (red) and HSA (blue) spectra.

Figure 4 shows a comparison of the spectra of pure HSA and GHSA. Separate gray zones show the areas in which the maximum differences between the spectra are concentrated. The band assignment is presented in Table 1.

Table 1. Band assignment for the HSA and GHSA spectra.

Band Position, cm^{-1}	Band Assignment	Band Position, cm^{-1}	Band Assignment
335, 414, 512	S-S	1107, 1130	rNH3, Lys
630	Tyr	1178, 1214	Tyr
650	ν C−S, Cys	1230–1300	Amid III
757	ρ(CH2)	1345	ω CH2, Trp
834, 859	Tyr	1455	δCH2 δCH3
907	ν(C–C)	1591	Phe, Trp
952	ν(C–C) (Random), Trp	1612	Tyr
1010, 1038	Phe	1662	Amid I (Random)

The bands were assigned according to the literature [38,39]. The main bands in both spectra are represented by vibrations at 512, 952, 1010, 1345, 1455, and 1662 cm^{-1}, which correspond to vibrations S-S, ν(C-C) in Trp, Phe, CH2 in Trp, δ(CH2), δ(CH3), and Amide I. The main contribution to the difference between the HSA and GHSA spectra is made by the bands at 512, 1345, 1445, 1591, and 1662 cm^{-1}. In general, we see a complex band pattern. Therefore, mathematical methods need to be used to accurately determine the differences in the spectra of proteins and their mixtures. Of course, for pure solutions of HSA and GHSA, the differences between the spectra are quite easy to notice, as can be seen in Figure 4. However, difficulties should be expected when working with mixtures when a rather small amount of GHSA is present in the main HSA solution.

3.3. Data Processing

Figure 5 depicts the preprocessed spectra obtained from the experiment.

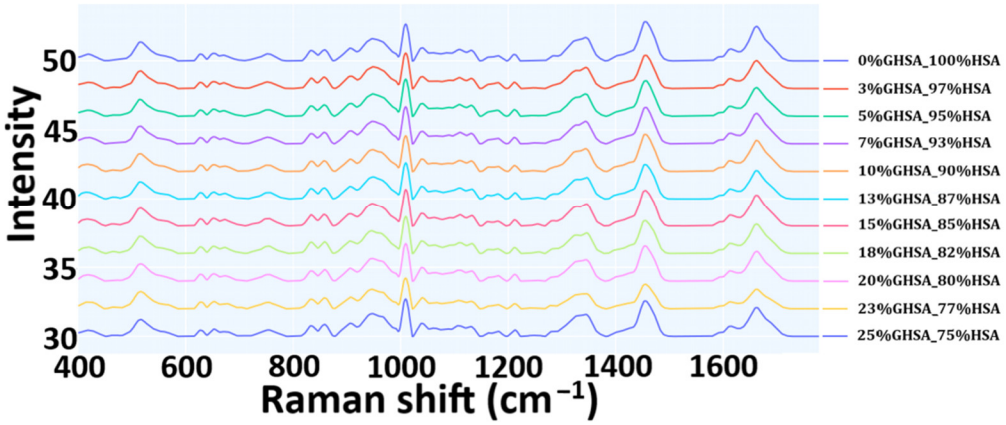

Figure 5. SERS spectra for mixtures HSA–GHSA.

As Figure 5 shows, all spectra are visually very similar; however, the main features characteristic of this group of spectra retained their shape after preprocessing. As suggested above, the gradual appending of GHSA in HSA at biologically significant concentrations changes the spectrum very insignificantly. With a simple visual comparison of the spectra, the changes are not obvious. Therefore, the use of special mathematical processing methods is necessary.

3.4. Classification Using LDA

A confusion matrix in Figure 6 shows the number of correct and incorrect predictions made by the classification model compared with the actual concentrations in the test dataset.

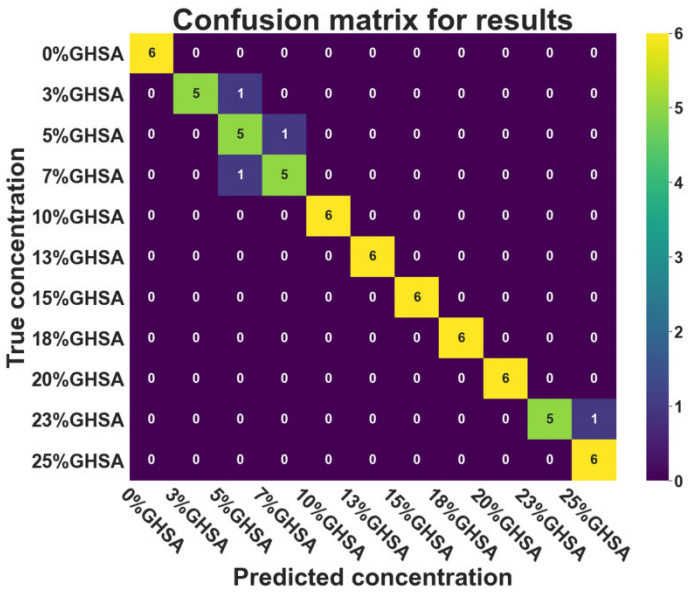

Figure 6. A confusion matrix. Each unit on diagonal represents correct predictions broken down by each concentration.

The classification model successfully separated the spectra of mixtures and made errors only when GHSA concentrations were close to each other (3–7%). It can be seen that if the model is wrong on the concentrations in the training set, then this predicted value is in close proximity (one unit on the grid) to the true value. Table 2 shows these results with the calculated quality metrics.

Table 2. The result of the classification.

Concentration GHSA	Precision	Recall	F1	Quantity
0% GHSA	1.00	1.00	1.00	6
3% GHSA	1.00	0.83	0.91	6
5% GHSA	0.71	0.83	0.77	6
7% GHSA	0.83	0.83	0.83	6
10% GHSA	1.00	1.00	1.00	6
13% GHSA	1.00	1.00	1.00	6
15% GHSA	1.00	1.00	1.00	6
18% GHSA	1.00	1.00	1.00	6
20% GHSA	1.00	1.00	1.00	6
23% GHSA	1.00	0.83	0.91	6
25% GHSA	0.86	1.00	0.92	6
Macro AVG	0.95	0.94	0.94	66

From the above, we can see that this problem has been successfully solved, achieving high precision and recall values (>90%). However, the classification model is poorly applicable to determining concentrations that were not in the training set. Therefore, we built a regression model based on the entire GHSA range.

3.5. Regression with Regularization

To predict the proportion of glycated albumin in the solution, we constructed a linear regression model with a regularization mechanism for a more robust model. To obtain such a model, we used regularization methods Lasso (L1) for feature selection and Ridge (L2) for weight optimization. Hyperparameters for this technique were selected using a grid

search algorithm with cross-validation for five folds on a training dataset. Figure 7 shows the results of the model's application to experimental data.

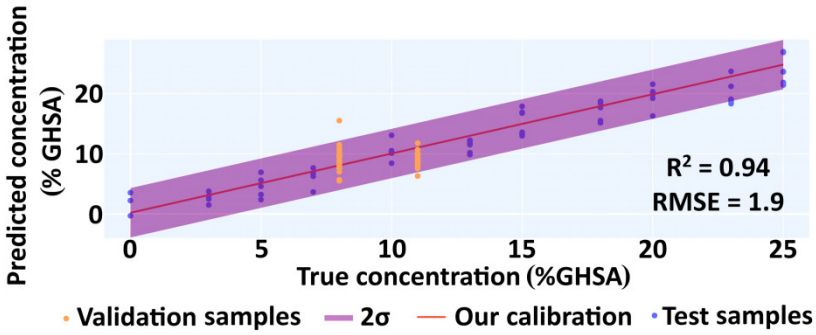

Figure 7. The predicted validation concentrations.

The average RMSE in determining the GHSA level in a test set was 1.9, which is less than 10% of the relative error. The predicted validation concentrations (8% and 11% of GHSA) fell into the double sigma interval (95%) from the test blind samples, which evinces a good generalization ability of the method due to the choice of the best parameters on cross-validation.

3.6. Loadings

The linearity of our calibration model allows us to visualize its parameters in the form of loading shifts, where each parameter corresponds to a contribution to the final result. Figure 8 demonstrates loading shifts that provide the maximum effect on the distinction of all spectra in the experiment.

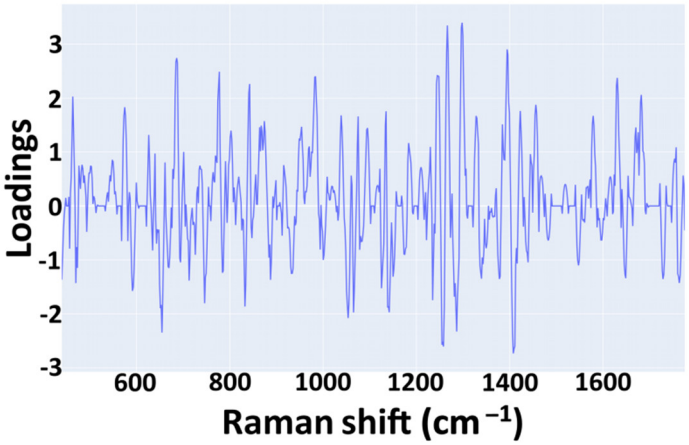

Figure 8. Loading shifts.

Analyzing the change in the spectra, we see that it is impossible to isolate the sequential evolution of one or more bands with an increase in the concentration of GHSA in HSA, which can be used to track the change in percentage. Rather, we observe a complete change in the spectra. Figure 8 shows that the main differences of some vibrational bands are located in the region of 650, 685, 757, 1260, 1268, 1297, 1408, 1631, and 1680 cm^{-1}. These peaks correspond to vibration bands ν(C-S) in Cys, (C-S), ρ(CH2), Amid III, Amid III, Amid III, unknown, Amid I, and Amid I. The observed change in the vibration bands for structurally alike analytes (such as GHSA and HSA) can be associated with a change in the structure of the micelle formed when proteins are dissolved in water upon combining different protein molecules in mixtures, depending on their concentration ratio. The rearrangement of molecule charges on the micelle surface likely leads to a change in the landing of molecules on the SERS substrate, as a result of which the SERS spectrum of

the mixture can change depending on the concentration of analytes in the mixture. In fact, HSA and GHSA are structurally the same substance containing different amounts of glucose, which is glycation. Consequently, the process of the appearance of the SERS spectrum from glycated molecules can be twofold. It can arise primarily due to a change in the chemical composition of the molecule due to the appearance of new chemical bonds in the molecule itself. On the other hand, this can lead to a change in the spectrum due to the alternative landing of the molecule on the substrate surface due to a change in the charge distribution on the surface of the molecule. The main thing here is that the substrate should be reproducible, ensuring the structural constancy of the adsorbed molecule on the surface. Since the size of the protein molecule is quite large, in particular, the effective size of the albumin molecule can be about 1 nm [40], which is comparable to the size of the region in which the SERS effect occurs [41], in this case, it is critically important for us to obtain morphologically reproducible substrates, which provides the proposed method for obtaining substrates and the method of analysis implemented with their help.

3.7. Error Analysis

The use of data augmentation made it possible to improve the values of quality metrics. Additionally, experiments were made using such popular algorithms as SVM and RF for comparison with our method [29]. The error analysis of the above algorithms was carried out by repeatedly sampling the training set, optimizing model parameters with fixed hyperparameters on it, and checking the corresponding metrics on the remaining part of the dataset. The results are shown as boxplots in Figure 9. Each plot shows the average value of the metric and its variance averaged over various dataset splits on the training and test samples, both with and without data augmentation.

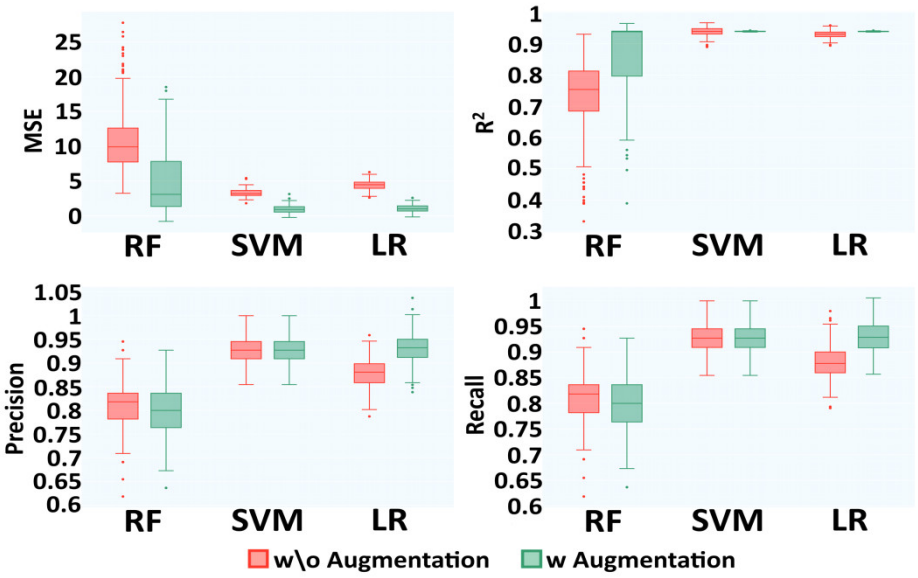

Figure 9. Error analysis and comparison.

Thus, the models we use show comparable or superior results in comparison with the reference ones. Since our regression model uses regularization, it is more stable due to the fact that not all the initial features are used to describe the target dependence. Additionally, the data augmentation method made it possible to reduce the error variance of our models and, in some cases, to improve the metrics values, as manifested in the reduced error interval and the shifted average.

4. Conclusions

In this work, we showed for the first time the possibility of quantitative determination of GHSA in a mixture with HSA by analyzing their SERS spectra in drying drops of

solutions without additional sample preparation. The obtained results have competitively advantageous values of MSE metrics in comparison with the widely used random forest and support vector machine; however, the regularization parameters enabled our model to surpass those approaches in determining the blind GHSA concentrations in solutions (8% and 11% GHSA of total solution concentration). The vibration bands that contribute to the differences between the spectra of samples with different GHSA concentrations were determined as follows: 650, 685, 757, 1260, 1268, 1297, 1408, 1631, and 1680 cm^{-1}. Based on these results, we can say that the method is simple and fast, and it can be used as the manner for medical measurement systems.

Author Contributions: Conceptualization, I.A.R. and I.N.K.; methodology, I.A.B., M.V.S., and R.R.S.; software, R.R.S.; validation, R.R.S., E.A.S., and N.L.N.; formal analysis, A.M.M.; investigation, E.A.S., R.R.S., and N.L.N.; resources, K.N.A. and I.A.R.; data curation, M.V.S., N.L.N. and I.N.K.; writing—original draft preparation, E.A.S. and R.R.S.; writing—review and editing, R.R.S., E.A.S., N.L.N., and I.A.B.; visualization, R.R.S., E.A.S., I.A.B., and M.V.S.; supervision, I.A.R., A.M.M., A.N.L., and I.N.K.; project administration, A.N.L. and A.M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Institute for Theoretical and Applied Electromagnetics RAS [Registration Theme FFUR-2024-0003]. This work was performed at the Unique Scientific Facility "Nanolayer".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: https://github.com/roberts2510/HSA-detection (accessed on 4 December 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Das, G.; Gentile, F.; Coluccio, M.L.; Perri, A.M.; Nicastri, A.; Mecarini, F.; Cojoc, G.; Candeloro, P.; Liberale, C.; De Angelis, F.; et al. Principal Component Analysis Based Methodology to Distinguish Protein SERS Spectra. *J. Mol. Struct.* **2011**, *993*, 500–505. [CrossRef]
2. Boginskaya, I.; Sedova, M.; Baburin, A.; Afanas'ev, K.; Zverev, A.; Echeistov, V.; Ryzhkov, V.; Rodionov, I.; Tonanaiskii, B.; Ryzhikov, I.; et al. SERS-Active Substrates Nanoengineering Based on e-Beam Evaporated Self-Assembled Silver Films. *Appl. Sci.* **2019**, *9*, 3988. [CrossRef]
3. Sigolaeva, L.V.; Nechaeva, N.L.; Ignatov, A.I.; Filatova, L.Y.; Sharifullin, T.Z.; Eichhorn, J.; Schacher, F.H.; Pergushov, D.V.; Merzlikin, A.M.; Kurochkin, I.N. In Situ SERS Sensing by a Laser-Induced Aggregation of Silver Nanoparticles Templated on a Thermoresponsive Polymer. *Biosensors* **2022**, *12*, 628. [CrossRef] [PubMed]
4. Podoynitsyn, S.N.; Sorokina, O.N.; Nechaeva, N.L.; Yanovich, S.V.; Kurochkin, I.N. Surface-Enhanced Raman Spectroscopy in Tandem with a Gradient Electric Field from 4-Mercaptophenylboronic Acid on Silver Nanoparticles. *Microchim. Acta* **2020**, *187*, 566. [CrossRef] [PubMed]
5. Lane, L.A.; Qian, X.; Nie, S. SERS Nanoparticles in Medicine: From Label-Free Detection to Spectroscopic Tagging. *Chem. Rev.* **2015**, *115*, 10489–10529. [CrossRef]
6. Schlücker, S. SERS Microscopy: Nanoparticle Probes and Biomedical Applications. *ChemPhysChem* **2009**, *10*, 1344–1354. [CrossRef]
7. David, C.; Guillot, N.; Shen, H.; Toury, T.; Chapelle, M.L. de la SERS Detection of Biomolecules Using Lithographed Nanoparticles towards a Reproducible SERS Biosensor. *Nanotechnology* **2010**, *21*, 475501. [CrossRef]
8. Kurochkin, I.N.; Eremenko, A.V.; Evtushenko, E.G.; Nechaeva, N.L.; Durmanov, N.N.; Guliev, R.R.; Ryzhikov, I.A.; Boginskaya, I.A.; Sarychev, A.K.; Ivanov, A.V.; et al. SERS for Bacteria, Viruses, and Protein Biosensing. In *Macro, Micro, and Nano-Biosensors*; Springer International Publishing: Cham, Switzerland, 2021; pp. 75–94.
9. Petti, L.; Capasso, R.; Rippa, M.; Pannico, M.; La Manna, P.; Peluso, G.; Calarco, A.; Bobeico, E.; Musto, P. A Plasmonic Nanostructure Fabricated by Electron Beam Lithography as a Sensitive and Highly Homogeneous SERS Substrate for Bio-Sensing Applications. *Vib. Spectrosc.* **2016**, *82*, 22–30. [CrossRef]
10. Suresh, V.; Ding, L.; Chew, A.B.; Yap, F.L. Fabrication of Large-Area Flexible SERS Substrates by Nanoimprint Lithography. *ACS Appl. Nano Mater.* **2018**, *1*, 886–893. [CrossRef]
11. Green, M.; Liu, F.M. SERS Substrates Fabricated by Island Lithography: The Silver/Pyridine System. *J. Phys. Chem. B* **2003**, *107*, 13015–13021. [CrossRef]

12. Coluccio, M.L.; Das, G.; Mecarini, F.; Gentile, F.; Pujia, A.; Bava, L.; Tallerico, R.; Candeloro, P.; Liberale, C.; De Angelis, F.; et al. Silver-Based Surface Enhanced Raman Scattering (SERS) Substrate Fabrication Using Nanolithography and Site Selective Electroless Deposition. *Microelectron. Eng.* **2009**, *86*, 1085–1088. [CrossRef]
13. Boginskaya, I.; Nechaeva, N.; Tikhomirova, V.; Kryukova, O.; Evdokimov, V.; Bulaeva, N.; Golukhova, E.; Ryzhikov, I.; Kost, O.; Afanasev, K.; et al. Human Angiotensin I-converting Enzyme Study by Surface-enhanced Raman Spectroscopy. *J. Raman Spectrosc.* **2021**, *52*, 1529–1539. [CrossRef]
14. Boginskaya, I.; Safiullin, R.; Tikhomirova, V.; Kryukova, O.; Nechaeva, N.; Bulaeva, N.; Golukhova, E.; Ryzhikov, I.; Kost, O.; Afanasev, K.; et al. Human Angiotensin I-Converting Enzyme Produced by Different Cells: Classification of the SERS Spectra with Linear Discriminant Analysis. *Biomedicines* **2022**, *10*, 1389. [CrossRef]
15. Baburin, A.S.; Gritchenko, A.S.; Orlikovsky, N.A.; Dobronosova, A.A.; Rodionov, I.A.; Balykin, V.I.; Melentiev, P.N. State-of-the-Art Plasmonic Crystals for Molecules Fluorescence Detection. *Opt. Mater. Express* **2019**, *9*, 1173. [CrossRef]
16. Yankovskii, G.M.; Komarov, A.V.; Puz'ko, R.S.; Baryshev, A.V.; Afanas'ev, K.N.; Boginskaya, I.A.; Bykov, I.V.; Merzlikin, A.M.; Rodionov, I.A.; Ryzhikov, I.A. Structural and Optical Properties of Single and Bilayer Silver and Gold Films. *Phys. Solid State* **2016**, *58*, 2503–2510. [CrossRef]
17. Baburin, A.S.; Ivanov, A.; Trofimov, I.; Dobronosovaa, A.; Melentiev, P.; Balykin, V.; Moskalev, D.; Pishchimova, A.; Ganieva, L.; Ryzhikov, I.; et al. Highly Directional Plasmonic Nanolaser Based on High-Performance Noble Metal Film Photonic Crystal. In Proceedings of the Nanophotonics VII, Strasbourg, France, 22–26 April 2018; Andrews, D.L., Nunzi, J.-M., Ostendorf, A., Bain, A.J., Eds.; SPIE: Bellingham, WA, USA, 2018; p. 159.
18. Baburin, A.S.; Ivanov, A.I.; Ryzhikov, I.A.; Trofimov, I.V.; Gabidullin, A.R.; Moskalev, D.O.; Panfilov, Y.V.; Rodionov, I.A. Crystalline Structure Dependence on Optical Properties of Silver Thin Film over Time. In Proceedings of the 2017 Progress In Electromagnetics Research Symposium—Spring (PIERS), St. Petersburg, Russia, 22–25 May 2017; pp. 1497–1502.
19. Koga, M.; Murai, J.; Saito, H.; Kasayama, S. Glycated Albumin and Glycated Hemoglobin Are Influenced Differently by Endogenous Insulin Secretion in Patients with Type 2 Diabetes. *Diabetes Care* **2010**, *33*, 270–272. [CrossRef]
20. Kohzuma, T.; Koga, M. Lucica^{TM} GA-L Glycated Albumin Assay Kit: A New Diagnostic Test for Diabetes Mellitus. *Mol. Diagnosis Ther.* **2010**, *14*, 49–51. [CrossRef]
21. Kouzuma, T.; Usami, T.; Yamakoshi, M.; Takahashi, M.; Imamura, S. An Enzymatic Method for the Measurement of Glycated Albumin in Biological Samples. *Clin. Chim. Acta* **2002**, *324*, 61–71. [CrossRef]
22. Roohk, H.V.; Zaidi, A.R. A Review of Glycated Albumin as an Intermediate Glycation Index for Controlling Diabetes. *J. Diabetes Sci. Technol.* **2008**, *2*, 1114–1121. [CrossRef]
23. Anguizola, J.; Matsuda, R.; Barnaby, O.S.; Hoy, K.S.; Wa, C.; DeBolt, E.; Koke, M.; Hage, D.S. Review: Glycation of Human Serum Albumin. *Clin. Chim. Acta* **2013**, *425*, 64–76. [CrossRef]
24. Kohzuma, T.; Yamamoto, T.; Uematsu, Y.; Shihabi, Z.K.; Freedman, B.I. Basic Performance of an Enzymatic Method for Glycated Albumin and Reference Range Determination. *J. Diabetes Sci. Technol.* **2011**, *5*, 1455–1462. [CrossRef] [PubMed]
25. Dingari, N.C.; Horowitz, G.L.; Kang, J.W.; Dasari, R.R.; Barman, I. Raman Spectroscopy Provides a Powerful Diagnostic Tool for Accurate Determination of Albumin Glycation. *PLoS ONE* **2012**, *7*, e32406. [CrossRef] [PubMed]
26. Nechaeva, N.L.; Boginskaya, I.A.; Ivanov, A.V.; Sarychev, A.K.; Eremenko, A.V.; Ryzhikov, I.A.; Lagarkov, A.N.; Kurochkin, I.N. Multiscale Flaked Silver SERS-Substrate for Glycated Human Albumin Biosensing. *Anal. Chim. Acta* **2020**, *1100*, 250–257. [CrossRef] [PubMed]
27. de Medeiros, A.D.; da Silva, L.J.; Ribeiro, J.P.O.; Ferreira, K.C.; Rosas, J.T.F.; Santos, A.A.; da Silva, C.B. Machine Learning for Seed Quality Classification: An Advanced Approach Using Merger Data from FT-NIR Spectroscopy and X-ray Imaging. *Sensors* **2020**, *20*, 4319. [CrossRef] [PubMed]
28. Lee, S.; Choi, H.; Cha, K.; Chung, H. Random Forest as a Potential Multivariate Method for Near-Infrared (NIR) Spectroscopic Analysis of Complex Mixture Samples: Gasoline and Naphtha. *Microchem. J.* **2013**, *110*, 739–748. [CrossRef]
29. Mekonnen, B.K.; Yang, W.; Hsieh, T.-H.; Liaw, S.-K.; Yang, F.-L. Accurate Prediction of Glucose Concentration and Identification of Major Contributing Features from Hardly Distinguishable Near-Infrared Spectroscopy. *Biomed. Signal Process. Control* **2020**, *59*, 101923. [CrossRef]
30. Arslan, M.; Guzel, M.; Demirci, M.; Ozdemir, S. SMOTE and Gaussian Noise Based Sensor Data Augmentation. In Proceedings of the 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, 11–15 September 2019; pp. 1–5.
31. Zou, H.; Hastie, T. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
32. Shvalya, V.; Filipič, G.; Zavašnik, J.; Abdulhalim, I.; Cvelbar, U. Surface-Enhanced Raman Spectroscopy for Chemical and Biological Sensing Using Nanoplasmonics: The Relevance of Interparticle Spacing and Surface Morphology. *Appl. Phys. Rev.* **2020**, *7*, 031307. [CrossRef]
33. Ducourtieux, S.; Podolskiy, V.A.; Grésillon, S.; Buil, S.; Berini, B.; Gadenne, P.; Boccara, A.C.; Rivoal, J.C.; Bragg, W.D.; Banerjee, K.; et al. Near-Field Optical Studies of Semicontinuous Metal Films. *Phys. Rev. B Condens. Matter Mater. Phys.* **2001**, *64*, 165403. [CrossRef]

34. Grésillon, S.; Aigouy, L.; Boccara, A.C.; Rivoal, J.C.; Quelin, X.; Desmarest, C.; Gadenne, P.; Shubin, V.A.; Sarychev, A.K.; Shalaev, V.M. Experimental Observation of Localized Optical Excitations in Random Metal-Dielectric Films. *Phys. Rev. Lett.* **1999**, *82*, 4520–4523. [CrossRef]
35. Breit, M.; Podolskiy, V.A.; Grésillon, S.; von Plessen, G.; Feldmann, J.; Rivoal, J.C.; Gadenne, P.; Sarychev, A.K.; Shalaev, V.M. Experimental Observation of Percolation-Enhanced Nonlinear Light Scattering from Semicontinuous Metal Films. *Phys. Rev. B* **2001**, *64*, 125106. [CrossRef]
36. Seal, K.; Nelson, M.A.; Ying, Z.C.; Genov, D.A.; Sarychev, A.K.; Shalaev, V.M. Growth, Morphology, and Optical and Electrical Properties of Semicontinuous Metallic Films. *Phys. Rev. B* **2003**, *67*, 035318. [CrossRef]
37. Brouers, F.; Blacher, S.; Lagarkov, A.N.; Sarychev, A.K.; Gadenne, P.; Shalaev, V.M. Theory of Giant Raman Scattering from Semicontinuous Metal Films. *Phys. Rev. B* **1997**, *55*, 13234–13245. [CrossRef]
38. Synytsya, A.; Alexa, P.; de Boer, J.; Loewe, M.; Moosburger, M.; Würkner, M.; Volka, K. Raman Spectroscopic Study of Serum Albumins: An Effect of Proton- and γ-Irradiation. *J. Raman Spectrosc.* **2007**, *38*, 1646–1655. [CrossRef]
39. Jurasekova, Z.; Marconi, G.; Sanchez-Cortes, S.; Torreggiani, A. Spectroscopic and Molecular Modeling Studies on the Binding of the Flavonoid Luteolin and Human Serum Albumin. *Biopolymers* **2009**, *91*, 917–927. [CrossRef]
40. Kiselev, M.A.; Gryzunov, I.A.; Dobretsov, G.E.; Komarova, M.N. Size of a Human Serum Albumin Molecule in Solution. *Biofizika* **2001**, *46*, 423–427.
41. Pilot, R.; Signorini, R.; Durante, C.; Orian, L.; Bhamidipati, M.; Fabris, L. A Review on Surface-Enhanced Raman Scattering. *Biosensors* **2019**, *9*, 57. [CrossRef]