



Article

Application of computational intelligence methods for the automated identification of paper-ink samples based on LIBS

Krzysztof Rzecki ¹, Tomasz Sośnicki ¹, Mateusz Baran ¹, Michał Niedźwiecki ¹,
Małgorzata Król ², Tomasz Łojewski ³, U Rajendra Acharya ^{4,5,6}, Özal Yildirim ⁷, and Paweł
Pławiak ^{1*}

¹ Faculty of Physics, Mathematics and Computer Science, Cracow University of Technology, Warszawska 24, 31-155 Krakow, Poland; krz@pk.edu.pl (K.Rz.); tsosnicki@pk.edu.pl (T.S.); plawiak@pk.edu.pl (P.P.); mbaran@pk.edu.pl (M.B.); nkg@pk.edu.pl (M.N.)

² Laboratory for Forensic Chemistry, Faculty of Chemistry, Jagiellonian University, Ingardena 3, 30-060 Krakow, Poland; krolm@chemia.uj.edu.pl (M.K.)

³ Faculty of Materials Science and Ceramics, AGH University of Science and Technology, Mickiewicza 30 Av., Krakow 30-059, Poland; lojewski@agh.edu.pl (T.Ł.)

⁴ Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore; Rajendra_Udyavara_ACHARYA@np.edu.sg (R.A.)

⁵ Department of Biomedical Engineering, School of Science and Technology, Singapore School of Social Sciences, Singapore; Rajendra_Udyavara_ACHARYA@np.edu.sg (R.A.)

⁶ School of Medicine, Faculty of Health and Medical Sciences, Taylor's University, 47500 Subang Jaya, Malaysia; Rajendra_Udyavara_ACHARYA@np.edu.sg (R.A.)

⁷ Department of Computer Engineering, Munzur University, Tunceli, Turkey; oyildirim@munzur.edu.tr (O.Y.)

* Correspondence: plawiak@pk.edu.pl (P.P.), krz@pk.edu.pl (K.Rz.)

Abstract: Laser-induced breakdown spectroscopy (LIBS) is an important analysis technique with applications in many industrial branches and fields of scientific research. Nowadays, the advantages of LIBS are impaired by the main drawback in the analysis of collected data. This procedure is essentially based on the comparison of lines present in the spectrum with a literature database. This paper proposes the use of various computational intelligence methods to develop a reliable and fast classification of non-destructively acquired LIBS spectra into a set of predefined classes. We focus on a specific problem of classification of paper-ink samples into 30 separate, predefined classes. For each of 30 classes (10 pens of each of 5 ink types combined with 10 sheets of 5 paper types plus empty pages) 100 LIBS spectra are collected. Four variants of preprocessing, seven classifiers (Decision trees, Random forest, *k*-Nearest Neighbour, Support Vector Machine, Probabilistic Neural Network, Multi-Layer Perceptron, and Generalized Regression Neural Network), 5-fold stratified cross-validation and test on an independent set (for methods evaluation) scenarios are employed. Our developed system yielded an accuracy of 99.08% with average classification time of about 0.12 s is obtained using the random forest classifier. Our results clearly demonstrates that machine learning methods can be used to identify the paper-ink samples based on LIBS reliably at a faster rate.

Keywords: classification; computational intelligence methods; discrimination power; LIBS; machine learning; paper-ink analysis

1. Introduction

During the decade, there have been important developments in laser induced breakdown spectroscopy (LIBS). This atomic emission spectroscopy technique, also known as laser induced plasma spectroscopy (LIPS), is used for qualitative and quantitative chemical analysis of samples in all states of matter [1]. In this technique, high-power, short-duration laser pulse causes an ablation of the analyzed material, which due to its high temperature (10 000K) dissociates into excited ions and atoms. When

24 plasma cloud cools down this excited species revert to lower energy states and emit optical radiation
25 which could be recorded and analyzed, revealing information about the elemental composition of
26 the sample [2]. LIBS spectra are generally very rich in emission lines coming from excited atoms and
27 ions occurring in a high temperature plasma cloud. Physical and chemical phenomena behind LIBS
28 are not fully understood yet, as they are very complex in nature. Nevertheless, LIBS applications
29 are recently rapidly growing due to the number of advantages of this method. The most important
30 ones are minimally destructive measurement with little or no sample preparation, efficiency and
31 possibility to analyze in real-time all elements in a single laser shot in all three states [3]. For solids both
32 mapping (2D) and depth profiling (3D) can be obtained. LIBS can be used for quantitative chemical
33 analysis, material identification and discrimination. This method can be applied in the laboratories and
34 industrial plants, and even at stand-off distances of tens of meters [4]. These properties predispose the
35 LIBS technique for use in many fields: space exploration [5], remote analysis of hazardous materials [6],
36 on-line quality control in various industries [7], cultural heritage studies [8], forensic chemistry [9,10],
37 geology [11], weld quality assurance [12], robotics [13] and many others [14–17].

38 Although currently LIBS is already an established technology, spectrochemical LIBS analysis is
39 not straightforward. Identification of elemental constituents in the sample is usually based on the
40 strongest lines present in the LIBS spectrum (so called persistent lines), which are compared with a
41 literature database collected for all the elements from the periodic table. This analysis is cumbersome
42 and time-consuming because the emission spectrum is determined also by the properties of the plasma,
43 not only by the composition of the examined sample [18].

44 At present, the great advantages of LIBS seem to be impaired to some extent by the main
45 drawbacks: problems coming from often poor signal reproducibility, the impact of sample composition
46 on signals recorded for individual components (matrix effects) and the difficulty of performing an
47 overall reliable data analysis (sometimes over 500 spectral lines need to be interpreted). Thus, a
48 question immediately raises whether the LIBS method can be supported by the modern achievements
49 in the field of computational intelligence [19] in an attempt to overcome the limitations of the current
50 LIBS spectrum analysis methodology. The analogous problem of classification of one dimensional
51 series of data is well known in computer science. The methods developed for this problem have been
52 successfully applied in many areas [20–27]. So, application of machine learning methods to LIBS
53 spectra of samples may certainly give a strong impact for further developments and applications.

54 Motivated by the aforementioned arguments we decided to apply computational intelligence
55 methods to the classification of paper-ink samples for forensic purposes, based on LIBS spectra of the
56 samples. The problem of discrimination of different paper-ink samples has already been addressed
57 in previous studies based on the LIBS spectrum analysis. Trejos et al. showed [28] that the highest
58 discrimination power (*DP*) (96.4%) was obtained when comparisons were done qualitatively by spectral
59 overlap of the regions of interest (3 different emission lines monitored per element) and quantitatively
60 followed by pairwise comparisons (1 emission line per element) using ANOVA (analysis of variance).
61 Kula et al. [29] presented the LIBS method as a useful tool in qualitative elemental differentiation
62 of ink samples. They have obtained the discrimination power (*DP*) coefficient of 61%, 82% and 83%
63 for red, black and blue inks, respectively. Elsherbiny and Nassef [30] studied the dependence of the
64 obtained spectra of the obtained spectra of various black gel inks on the wavelength of laser excitation
65 reporting the *DP* in the range of 88% to 91%. In the next study [31], the elemental analysis with the
66 use of LIBS was performed on writing and inkjet inks, toners as well as on office paper. The LIBS
67 results supported by pairwise comparison analysis (ANOVA with Tukey's post hoc test) provided
68 discrimination power of 98.4% (3 indistinguishable/190 compared pairs) for the toners and 100% for
69 the inkjet inks. Moreover, these three undifferentiated toner pairs were discriminated using the Student
70 t-test at a 95% confidence limit. The authors claimed that LIBS as a suitable tool for the determination
71 of elemental composition of sample can be a part of a procedure for questioned document examination.

72 The problem of paper-inks classification, on the other hand, is less common in the literature. To the
73 best of authors knowledge there are only two articles dealing with LIBS spectra of paper-ink samples

Table 1. List of ballpoint pens used in the experiment.

Company	Model	Series	Sample ID
Bic	Orange fine blue	070330101593	B
Rystor	Kropka 0.5 1968	5907548914507	R
Staedtler	Stick 430F	4007817410660	S
Staedtler	Ball 432F	4007817432020	SB
Toma	Sunny fine 050	590113305113	T

74 employed in classification/fitting problems by means of computational intelligence techniques. In
 75 a paper by Hoehses et al. [32], the benefit of applying several independent chemometric methods to
 76 LIBS data was demonstrated. A consecutive methodology of applying soft independent modelling
 77 of class analogy (SIMCA) and partial least-squares discriminant analysis (PLS-DA) enabled the
 78 step-wise classification of data and separation of inks that were not identified by principal component
 79 analysis (PCA). The SVM yielded a correct classification rate of 87% and cross-validation accuracy
 80 amounted to 81%. In the second paper [33], multiple methods such as three comparative functions
 81 (linear correlation, overlapping integral, and sum of squared deviations) and two advanced statistical
 82 methods (multivariate curve resolution alternating least squares (MCR-ALS) with classification tree and
 83 discriminant analysis (DA)) were applied to statistically evaluate LIBS spectra. The newly introduced
 84 MCR-ALS/DA methodology showed identification of the paper and printer type with an accuracy of
 85 96.3% and 83.3%, respectively.

86 In the present paper, we show that the similar problem can be solved using computational
 87 intelligence methods. With our approach much better *DP* is achieved and the processing time is greatly
 88 reduced. Our results show that the machine learning algorithms can be used to analyze the LIBS
 89 spectra.

90 The classification problem discussed in this paper is very difficult due to similarities between
 91 LIBS spectra from different paper-ink samples and differences between such spectra from one single
 92 class. Additionally, we operate over the long samples with significant noise. To solve this classification
 93 problem, we tested many preprocessing ways and computational intelligence methods, however the
 94 paper presents only selected and best ones.

95 2. Materials and Methods

96 2.1. Materials

97 Fifty ballpoint pens (10 items of each of 5 models) produced by four different manufacturers from
 98 Germany and Poland (more details in Table 1) were purchased in Poland. Fifty sheets (10 sheets of
 99 each of 5 types) of five Canadian certified reference papers were used (papers denoted A, D, L, N and
 100 O from the set “Fillers in paper” supplied by A.S.O. Design Canada).

101 Each particular ink from each of 10 pens of 5 types was deposited (as straight lines) on each of 10
 102 sheets of 5 types of papers in the form of straight lines, using standard hand pressure. All 2 500 (50
 103 sheets of papers with 50 deposits each) paper-ink samples were placed in plastic bags and stored in
 104 darkness at room temperature.

105 In the experiment data for 30 classes (A, A+B, A+R, A+S, A+SB, A+T, D, D+B, . . . , O+SB, O+T)
 106 based on the combination of ink, paper and empty papers were recorded.

107 2.2. LIBS

108 The analysis of all paper-ink samples was carried out using a laser induced breakdown
 109 spectroscopy system LIBS-6 (Applied Photonics, United Kingdom). It consists of an integrated
 110 Q-switched Quantel Ultra Nd-YAG laser (Quantel, France) working at $\lambda = 1064$ nm emitting a
 111 maximum energy equal to 150 mJ due to one laser pulse (6 ns), and an Avaspec-2048-2-USB2 fibre optic

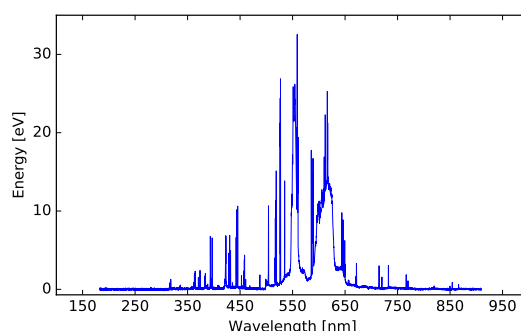


Figure 1. Visualization of a signal from the class A.

112 Czerny-Turner spectrometer (6-channel) with a CCD detector (Avantes, The Netherlands). The system
113 was also equipped with a camera enabled to observe the analyzed object and a movable sample table.
114 The LIBS-6 system was operated by LIBSoft V6.0.1 software (Applied Photonics, United Kingdom).
115 Under normal conditions, because of no moving elements inside, a wavelength calibration of the
116 spectrometer was not required. Every measurements were conducted in air under atmospheric pressure.

117 A Q-switch delay time of 165 μs , the integration delay time of 1.27 μs and integration time of
118 1.2 ms were used. Samples were analysed directly without any special preparation. They were placed
119 at the sample table at the focal point of the focusing lens (at a distance determined by the nozzle about
120 70 mm from the optical head). The diameter of the ablation spot, ranged from 0.6 mm to 1 mm, was
121 dependent on the analysed material.

122 The results of LIBS analysis are an emission spectrum – intensity distribution of the radiation
123 energy (expressed in electron volts) emitted by the analysed object depends on the wavelength
124 measured in nanometres. The emission spectra were collected in the UV-vis range (185 nm to 904 nm,
125 spectral resolution 0.1 nm).

126 In total, 30 classes of experimental data listed above were recorded. For each class 100 LIBS
127 spectra were acquired, resulting in 3 000 samples of the LIBS emission spectrum. A sample spectrum is
128 shown in Figure 1. The spectrum of each sample is available on our website [34].

129 2.3. Data Analysis

130 The analysis of the LIBS spectrum consisted of the following steps:

- 131 1. Independent preprocessing of LIBS samples.
- 132 2. Selection of data samples for the cross validation and testing sets.
- 133 3. Data analysis based on computational intelligence methods.
- 134 4. Evaluation of the results.

135 Steps of the experiment are presented in the flowchart in Figure 2.

136 2.3.1. Signal preprocessing

137 The initial preprocessing of the raw data was performed to reduce the number of data points
138 corresponding to individual samples of the LIBS spectrum. This preprocessing step was applied
139 because the spectral range of LIBS is very broad (11 746 spectral lines in this case) and from the
140 viewpoint of computational intelligence applied to classification tasks it contains irrelevant information.
141 It can be expected that the reduction of the datasets enhances extraction of characteristic features of the
142 distinctive classes (types of material) [35]. Four preprocessing steps were taken under consideration:

- 143 • Data reduction by removing the data points from the beginning (first 3 746 data points) and from
144 the end (last 1 000 data points) of the LIBS spectrum, which do not contain relevant information.
145 7 000 data points are left for further analysis.

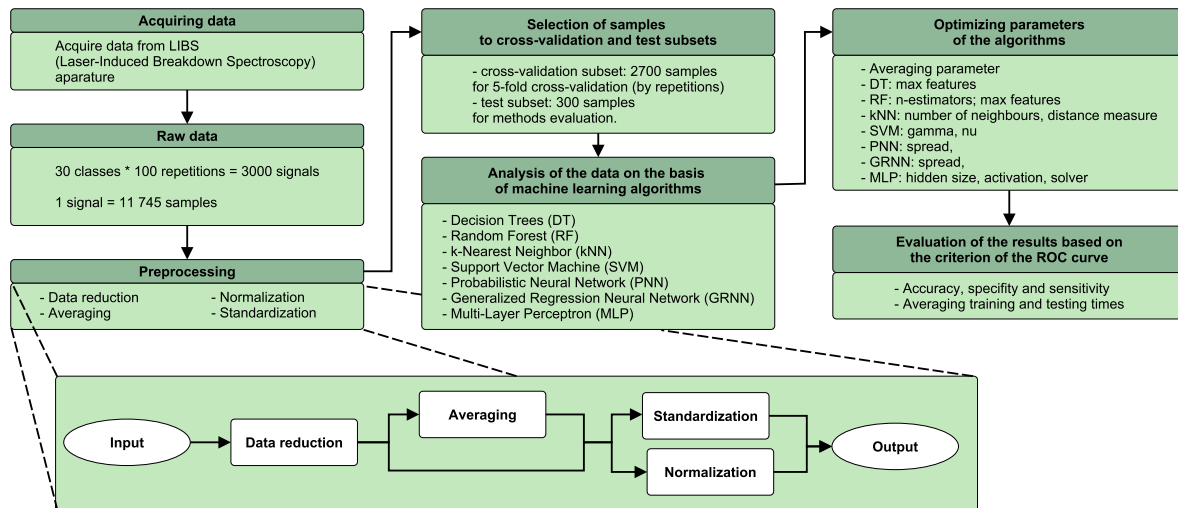


Figure 2. Flow chart of the experiment.

- 146 • Normalization of energy values to the interval $[0, 1]$.
- 147 • Standardization of energy values, so the mean value becomes equal to 0 and standard deviation
- 148 becomes equal to 1.
- 149 • Application of arithmetic averaging over consecutive number AP (called averaging parameter) of
- 150 data points after data reduction and prior to either normalization or standardization. As a result,
- 151 we reduce the original length of the data vector by a factor of AP . The tested vales of AP ranged
- 152 from 5 to 100.

153 Four preprocessing ways based on these steps were constructed and evaluated: data reduction,

154 averaging (applied or not), and either standardization or normalization. These steps are depicted at

155 the bottom of Figure 2.

156 Custom software in Python language was developed for data preprocessing tasks.

157 2.3.2. Cross-validation

158 The set of 3 000 LIBS emission spectrum samples was divided into two subsets [36] a training

159 subset containing 90% (2 700) of samples and a test subset containing 10% (300) of samples. Selection

160 of samples for both subsets was performed in a stratified way. The cross-validation subset was used

161 for parameter optimization and then the test subset was used for final evaluation of different types of

162 classifiers with optimized parameters.

163 The 5-fold stratified cross-validation method was applied to build the training and validation

164 data sets. In each of the five analyzed combinations a training set of 72 LIBS spectra from each of 30

165 classes (2 160 in total) and a validation set of 18 LIBS spectra from each of 30 classes (540 in total) were

166 used."

167 Evaluation of the classifiers with optimized parameters was performed separately for each training

168 subset using the test subset. Results of evaluation based on each training subset were obtained by

169 averaging the folds.

170 2.3.3. Computational Intelligence Methods

171 The samples of LIBS spectra, after preprocessing and cross-validation were fed to the classifiers:

172 Generalized Regression Neural Network (GRNN) [37], Probabilistic Neural Network (PNN) [38],

173 Multi-Layer Perceptron (MLP) [39], Support vector machine (SVM) [40], Decision trees (DT) [41],

174 k -nearest neighbour (kNN) classifier [42] and Random forest (RF) [41].

175 Each method is potentially dependent on a set of parameters that are either quantitative of

176 qualitative. They have an influence on the overall performance of the method. These parameters were

177 separately optimized for each machine learning algorithm to receive the lowest number of erroneous
 178 classes. The basic categorical parameters for each method were set to constants listed in Table 2.

Table 2. Computational intelligence methods and their basic parameters used for LIBS spectra identification.

No.	Method	Configuration
1	Decision trees	Criterion: gini, Splitter type: best, Maximum depth: none
2	Random forest	Criterion: gini, Maximum depth: none
3	kNN	Distance metric: Minkowski
4	SVM	Type: nuSVC, Type of kernel function: radial basis function
5	Neural Network	Type: PNN
6	Neural Network	Type: GRNN
7	Neural Network	Type: MLP

179 Methods based on adaptive neuro-fuzzy inference system (ANFIS) [43] and Gaussian process [44]
 180 were also tested, but they are computationally intensive and take long time to receive results even with
 181 reduced data set.

182 The custom software developed for this study uses the scikit-learn Python library [45] as a source
 183 of implementations of employed computational intelligence algorithms

184 2.3.4. Evaluation criteria

185 Evaluation of classification process was based on methodology described in [46,47]. The 5-fold
 186 cross-validation strategy was adopted in our experiment. Five performance parameters namely
 187 accuracy (*ACC*), sensitivity (*SEN*), specificity (*SPE*), mean values and Cohen's kappa (κ) were calculated.
 188 The mean values were calculated to estimate the overall performance of the computational intelligence
 189 methods used in this study separately for the task of recognition of each of the different classes of LIBS
 190 spectra.

191 To test whether there are classes which are classified with higher accuracy, specificity or sensitivity,
 192 we calculated for each class *S* the number of true positives $TP(S)$, false positives $FP(S)$, true negatives
 193 $TN(S)$ and false negatives $FN(S)$ separately. Then, the accuracy $ACC(S)$, sensitivity $SEN(S)$, and
 194 specificity $SPE(S)$ with respect to class *S* are defined as averages over five folds of cross-validation
 195 ($K = 5$ is the number of folds):

$$ACC(S) = \frac{1}{K} \sum_{i=1}^K \frac{TP_i(S) + TN_i(S)}{TP_i(S) + FP_i(S) + TN_i(S) + FN_i(S)}, \quad (1)$$

$$SEN(S) = \frac{1}{K} \sum_{i=1}^K \frac{TP_i(S)}{TP_i(S) + FN_i(S)}, \quad (2)$$

$$SPE(S) = \frac{1}{K} \sum_{i=1}^K \frac{TN_i(S)}{TN_i(S) + FP_i(S)}, \quad (3)$$

196 where $TP_i(S)$, $FP_i(S)$, $TN_i(S)$, $FN_i(S)$ are, respectively, the number of true positives, false positives,
 197 true negatives and false negatives for the *i*th fold of cross-validation with respect to class *S*, $i =$
 198 $1, 2, \dots, K$.

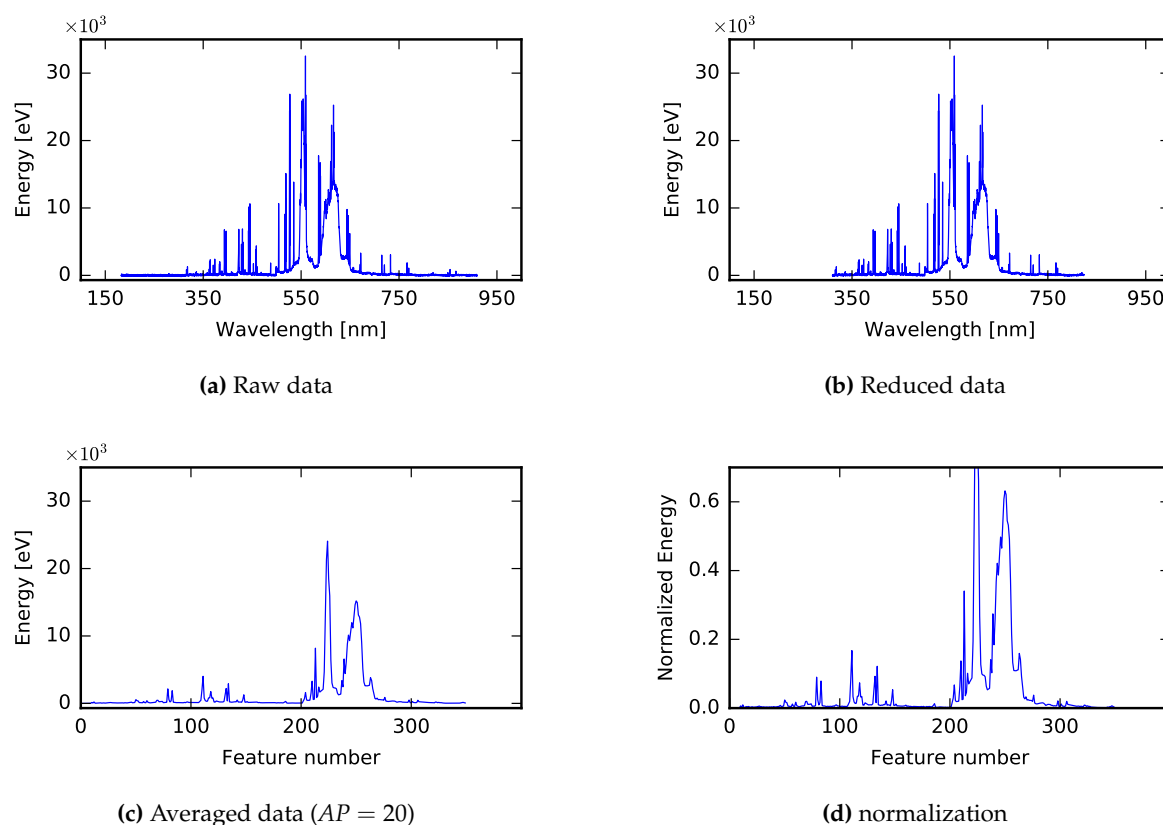


Figure 3. Visualization of data preprocessing for single sample from the class A. Figure (a) depicts an example of a raw spectrogram, (b) shows data after reduction, (c) represents averaged data for AP equal to 20, which is further normalized as shown in (d).

199 The overall values of ACC , SEN and SPE for the classification system are the arithmetic means of
 200 $ACC(S)$, $SEN(S)$ and $SPE(S)$ over all classes.

201 To evaluate the degree of discrimination between two different samples discrimination power
 202 coefficient (denoted DP) was proposed [48]. DP is the ratio of the number of correctly identified pairs
 203 of test samples (identified as from different classes) to the number of all possible pairs of test samples.
 204 It can be calculated by the following equation:

$$DP = \frac{2D}{T(T-1)} = 1 - \frac{2N}{T(T-1)}, \quad (4)$$

205

206 where:

- 207 • D – the number of differentiated pairs, that is the number of pairs correctly identified as belonging
- 208 to the same class or correctly identified as belonging to different classes,
- 209 • N – the number of non-differentiated pairs, $N = T(T-1)/2 - D$,
- 210 • T – the total number of analysed samples (the total number of possible pairs of samples is equal
- 211 to $T(T-1)/2$).

212 3. Results

213 An example of a raw spectrum and the output from successive stages of signal preprocessing
 214 (data reduction, averaging, normalization) are shown in Figure 3. The averaging stage is optional and
 215 normalization can be replaced by standardization.

216 The averaging within preprocessing is performed mainly to increase the speed of training and
217 testing, however it can also be described as a feature extraction procedure. Obviously the higher AP is,
218 the faster training and validation are for each of seven tested classification methods. The computation
219 time and classification accuracy were investigated for AP equal to 5, 10, 20, 50, 100, 150, 200 and
220 250. The best trade-off between the number of erroneous classifications and the reduction in the
221 computational time was achieved for AP equal to 20. For this value of AP , the training computation
222 times were shorter up to 266 times and the testing times were shorter up to 150 times relatively to the
223 processing without averaging (values of speed-up were dependent on the classification method). Thus,
224 for AP equal 20 the number of final spectral lines used decreased from 11 746 to 350.

225 Parameter selection is a key part in reaching the optimal overall performance of a classification
226 system. There are many possible options available, so it is important to analyze the outcomes of
227 experiments for different values of parameters to demonstrate the possibilities of machine learning
228 methods. Particular classification methods depend on the various basic parameters set as listed in
229 Table 2 and some other parameters that were optimized. Optimization of these parameters was
230 performed in two steps. The first step was to find out the general range of values for each parameter
231 for the fine-tuning procedure. Then the detailed grid search of the selected ranges of parameters was
232 performed.

233 The classification algorithms, the tuning parameters and range of these parameter values are
234 described below.

- 235 • Decision trees – the number of features to consider when looking for the best split was optimized in
236 two different ranges: from 100 to 7 000 with step equal to 100 when no averaging in preprocessing
237 was done, and from 10 to 350 with step equal to 10 when averaging in preprocessing was
238 performed.
- 239 • Random forest – two parameters dependent on averaging were optimized during preprocessing.
240 If averaging was not applied, the number of trees in the forest was optimized in range from 10 to
241 200 with step 10 and from 200 to 1 000 with step equal to 50 and the number of features to consider
242 when looking for the best split was optimized in the same range. If averaging was performed,
243 both parameters were optimized in range from 10 to 350 with step 20.
- 244 • k NN – the number of neighbours was optimized in the range from 1 to 4 and exponent used to
245 calculate the Minkowski distance was optimized from 1 to 10 with step 1.
- 246 • SVM – the gamma parameter of the RBF kernel function was optimized in range from 0.01 to 1.00
247 with step 0.01 and the nu parameter of the nu-SVC algorithm, related to the error tolerance of the
248 SVM classification, was optimized in range from 0.01 to 1.00 with step 0.01.
- 249 • PNN – the radius (the spread) of the kernel function of the network (standard deviation for the
250 probability density function of the normal distribution). This parameter was optimized in range
251 from 0.01 to 0.20 with step equal to 0.01 when normalization in preprocessing was used and from
252 0.1 to 1.0 with step 0.1 when standardization was used.
- 253 • GRNN – the spread parameter with identical meaning and range as in PNN was optimized;
- 254 • MLP – the number of neurons was optimized in a range from 10 to 200 with step equal to 10. The
255 activation function was selected from 'identity', 'logistic', 'tanh', 'relu'. The solver for weight
256 optimization was chosen from 'lbfgs' (an optimizer in the family of quasi-Newton methods), 'sgd'
257 (a stochastic gradient descent) or 'adam' (a stochastic gradient-based optimizer proposed in [49]).

258 These optimal parameters were selected in a series of experiments to maximize the mean ACC , SEN
259 and SPE values described in the section 2.3.4.

260 The best results of the experiments with different preprocessing methods, machine learning
261 algorithms, based on 5-fold cross-validation (3 000 samples were divided into training set containing
262 2 160 samples, validation set with 540 samples and test sets containing 300) are presented in Table
263 3. The preprocessing involves data reduction, with or without averaging (as in table noted) and
264 standardization. Generally the classification results were worse, when the preprocessing included
265 normalization instead of standardization. In the table, we show the values of ACC , SEN , SPE (and

Table 3. Performance results for optimized parameters.

Averaging	Method	Parameters	ACC (%)	SEN (%)	SPE (%)	MEAN (%)	κ (%)	DP (%)	Training time [s]	Testing time [s]
No	Decision trees	Max features = 6100	98.08	71.13	99.00	89.40	70.14	98.40	24.13	0.0002
No	Random forest	N estimators = 700 Max features = 950	99.08	86.27	99.53	94.96	85.79	99.08	1667.57	0.1198
Yes	kNN	N = 1 Exponent = 1.00	97.14	57.07	98.52	84.24	55.59	97.94	0.0505	0.0013
Yes	SVM	Nu = 0.13 Gamma = 0.02	98.92	83.80	99.44	94.05	83.24	99.00	3.60	0.0012
Yes	PNN	Spread = 0.9	97.25	58.73	98.58	84.85	57.31	97.98	0.0115	0.0194
Yes	GRNN	Spread = 0.1	96.94	54.07	98.42	83.14	52.48	97.56	0.0085	0.0070
No	MLP	N. of neurons = 120	98.22	73.27	99.08	90.19	72.34	98.36	221.76	0.0011

266 their mean value denoted *MEAN*), κ and *DP*, and average time of the training for all 5-fold training
 267 and testing samples.

268 Table 4 shows the values of *ACC(S)*, *SEN(S)*, and *SPE(S)* for all *S* classes and all classifiers with
 269 appropriate preprocessing methods. It can be seen from the table that the most easily distinguishable
 270 (highest *ACC(S)*, *SPE(S)* and *SEN(S)*) samples are in 6 classes: A, A+T, D, N, O and O+T (*ACC(S)*,
 271 *SPE(S)* and *SEN(S)* reached the value of 100%) using random forest method. Additionally, samples
 272 from 2 classes: L and N+T are well recognized by SVM and MLP methods as well. The least
 273 distinguishable classes are L+S and O+B with value of *SEN(S)* less than 50% by random forest
 274 method, however *SPE(S)* value was over 99%. It means that the samples from these classes are often
 275 assigned to other classes.

276 The computations were run on a virtual machine with Intel(R) Xeon(R) CPU L5640 @ 2.27 GHz
 277 (12 cores without HyperThreading were used) and 20 GB RAM DDR3 1333 MHz. Independent
 278 computations were performed in parallel. The total processing time of each step of computations
 279 depends on selected method and its parameters. The training time varied between less than a second
 280 for GRNN, PNN and kNN, and almost half an hour for random forest classifiers. The testing time for
 281 one sample varied between about 1-2 ms for decision tree and MLP, and more than 0.1 s for random
 282 forest, PNN and GRNN classifiers. The time necessary to complete the training was longer but it is
 283 less important than the classification speed that is critical at the testing stage. The time required for the
 284 identification of a particular sample was less than 1 second (however the exact value depends on the
 285 method) and is shown in Table 3.

286 4. Discussion

287 The results of experiments confirm that the computational intelligence methods can be used to
 288 analyze LIBS data and obtain accurate classification of paper-ink samples (please see Table 3). We
 289 have obtained an accuracy of 99.08% for the best classifier. The results of the present study can not be

Table 4. Results of classification per class.

Table with columns for Class, Averaging Standardization Decision tree, Random forest, kNN, SVM, PNN, GRNN, MLP and metrics ACC(%), SPE(%), SEN(%).

Table with columns for Class, No averaging Standardization Decision tree, Random forest, kNN, SVM, PNN, GRNN, MLP and metrics ACC(%), SPE(%), SEN(%).

290 directly compared with those obtained in [29,30] focusing on the problem of discrimination between
291 two samples which were not assigned to a priori classes. However, the methods used in [29,30] used
292 Plasus SpecLine 2.13 software for the identification of spectral lines and National Institute of Standards
293 and Technology (NIST) spectral database, gave values of the DP in the range from 85% to 92%. The
294 same standard methods with visual spectra was applied to the data collected in our study gave DP
295 equal to 90.6%. After adopting our algorithms to the problem of discrimination between two different
296 samples, we achieved the DP of 99.0% with SVM and more than 98% with decision trees, random forest,
297 PNN and MLP classifiers. The limitation of applying our solution to the problem of discrimination is
298 that the classes must be predefined.

299 Additionally, application of an automated method based on machine learning greatly reduced the
300 time required for spectrum recognition. The recognition time for a single spectrum was less than 1
301 second.

302 The detailed analysis of the Table 3 shows that the best method for the class of problems addressed
303 in the present study is the random forest classifier. Parameters for this classifier were optimized using
304 the 5-fold cross-validation using averaged and standardized data. The classifier during test over
305 independent test set was able to correctly recognize 1 294 cases of LIBS spectra out of 1 500 cases and

306 had no erroneous recognitions in training sets (10 800 cases). This classifier achieved an accuracy of
307 99.08%, sensitivity of 86.27%, specificity of 99.53%, kappa of 85.79% and *DP* of 99.08% for the test set.

308 Taking into account the number of erroneous classifications under different settings, the worst
309 classifier was GRNN, however it was the fastest classifier during training.

310 Four preprocessing methods were applied on the raw data prior to the classification. Hence, we
311 have shown that the success of classification depends on careful selection of preprocessing procedures.
312 However, although in most cases the actual selection of the pair preprocessing-classifier determines
313 the overall performance of the classification method, for some specific selections of classifier the final
314 classification performance depends only marginally on the preprocessing. For this reason, it is not
315 possible to select a single best preprocessing method.

316 The estimation of an optimal *AP* for decision trees, *k*NN, SVM and GRNN methods will be the
317 subject of the future research. It should be pointed out that the problem of optimal data reduction is
318 important. The length of data series influences the time needed to train the classifiers although it has
319 only marginal effect on the time needed for classification of a test sample.

320 Looking at the spectral data processed in the experiment, the averaging does not seem to be an
321 appropriate preprocessing method for data reduction because the differences between different kinds
322 of samples are inconspicuous. Authors took two other methods into consideration: resampling and
323 wavelet decomposition. The first method resamples the input data at given wavelengths of the original
324 sequence. To obtain 350 samples, resampling with $AP = 20$ was performed. This method, yielded
325 worse results during identification when averaging step was replaced. The second method performs
326 one dimensional Haar wavelet decomposition [50]. The initial results are promising, however this
327 method requires more detailed investigation in the future. This technique needs further research to
328 determine the type of wavelet decomposition structure and its parameters which will lead to the best
329 results.

330 The results of the present study show that modern computational intelligence methods can
331 be used for the classification of LIBS spectra into predefined classes, which solves a broad class of
332 problems related to LIBS applications. The main limitation of this study is that with the chosen settings
333 another broad class of LIBS applications, that is analysis of elemental composition, cannot be solved.
334 The problem of classification addressed in the present study is certainly simpler than the problem
335 of detailed analysis of the composition of materials. However, successful solution of this problem
336 encourages further research in this area. We suppose, because the problem of analysis of composition
337 of materials based on a LIBS spectrum may occur to be equivalent to the recognition of the presence of
338 specific spectral lines, this problem can be brought to a series of classification subproblems.

339 5. Conclusions

340 The aim of this study is to solve the difficult problem of distinguishing paper-ink samples. Hence,
341 we have designed a computational intelligence system to solve this problem using LIBS spectra.
342 Difficulty of this problem is caused by the high variability of spectra within a single class of paper-ink
343 samples and strong similarity of samples from different classes.

344 A machine learning system is developed and based on results presented in Table 3 and Table 4,
345 we can conclude that the system performed its task for majority of LIBS spectra in a short time. The
346 described method has sensitivity of 86% and it needs to be improved further. The main limitation of
347 this method is that we do not list particular elements of the tested material, however it also should be
348 taken under consideration in future work. Additional feature extraction and classification methods
349 can be applied in the future.

350 **Author Contributions:** The work presented in this paper was a collaboration of all authors. The manuscript was
351 revised by all authors. Conceptualization, M.K., K.Rz., T.S., and P.P.; Methodology, K.Rz., M.K., and P.P.; Software,
352 T.S.; Validation, K.Rz., and P.P.; Investigation, M.K., K.Rz., T.S., and P.P.; Resources, M.K., M.N., and T.L.; Data
353 Curation, M.K., T.S., and P.P.; Writing – Original Draft Preparation, M.K., K.Rz., and P.P.; Writing – Review &
354 Editing, M.B., K.Rz., O.Y., P.P., and R.A.; Visualization, T.S., and P.P.; Supervision, K.Rz., P.P., and R.A.;

355 **Funding:** This research received no external funding.

356 **Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design
357 of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the
358 decision to publish the results.

359

- 360 1. Singh, J.; Thakur, S. *Laser-Induced Breakdown Spectroscopy*, 1st ed.; Elsevier The Netherlands: UK, 2007.
- 361 2. Anabitarte, F.; Cobo, A.; Lopez-Higuera, J.M. Laser-Induced Breakdown Spectroscopy: Fundamentals,
362 Applications, and Challenges. *International Scholarly Research Notices Spectroscopy* **2012**, *2012*.
363 doi:10.5402/2012/285240.
- 364 3. Galbács, G. A critical review of recent progress in analytical laser-induced breakdown spectroscopy.
365 *Analytical and Bioanalytical Chemistry* **2015**, *407*, 7537–7562. doi:10.1007/s00216-015-8855-3.
- 366 4. Stelmazczyk, K.; Rohwetter, P.; Méjean, G.; Yu, J.; Salmon, E.; Kasparian, J.; Ackermann, R.; Wolf, J.P.;
367 Wöste, L. Long-distance remote laser-induced breakdown spectroscopy using filamentation in air. *Applied*
368 *Physics Letters* **2004**, *85*, 3977–3979. doi:http://dx.doi.org/10.1063/1.1812843.
- 369 5. Wiens, R.; Maurice, S. The ChemCam investigation: compositions at the curiosity rover landing site.
370 Geological Society of America Abstracts with Programs, 2012, Vol. 44, p. 190.
- 371 6. Ramirez-Cedeno, M.; Ortiz-Rivera, W.; Pacheco-Londono, L.; Hernandez-Rivera, S. Remote Detection of
372 Hazardous Liquids Concealed in Glass and Plastic Containers. *Sensors Journal, IEEE* **2010**, *10*, 693–698.
373 doi:10.1109/JSEN.2009.2036373.
- 374 7. Noyel, M.; Thomas, P.; Charpentier, P.; Thomas, A.; Brault, T. Implantation of an on-line quality process
375 monitoring. Industrial Engineering and Systems Management (IESM), Proceedings of 2013 International
376 Conference on, 2013, pp. 1–6.
- 377 8. Ortiz, R.; Ortiz, P.; Colao, F.; Fantoni, R.; Gómez-Morón, M.; Vázquez, M. Laser spectroscopy and imaging
378 applications for the study of cultural heritage murals. *Construction and Building Materials* **2015**, *98*, 35 – 43.
379 doi:10.1016/j.conbuildmat.2015.08.067.
- 380 9. Rinke-Kneapler, C.; Sigman, M. 15 - Applications of laser spectroscopy in forensic science. In
381 *Laser Spectroscopy for Sensing*; Baudalet, M., Ed.; Woodhead Publishing, 2014; pp. 461 – 495.
382 doi:10.1533/9780857098733.3.461.
- 383 10. Król, M.; Kowalska, D.; Kościelniak, P. Examination of Polish Identity Documents
384 by Laser-Induced Breakdown Spectroscopy. *Analytical Letters* **2018**, *51*, 1592–1604,
385 [<https://doi.org/10.1080/00032719.2017.1384833>]. doi:10.1080/00032719.2017.1384833.
- 386 11. Gaft, M.; Sapir-Sofer, I.; Modiano, H.; Stana, R. Laser induced breakdown spectroscopy for bulk minerals
387 online analyses. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2007**, *62*, 1496 – 1503. A Collection of Papers
388 Presented at the 4th International Conference on Laser Induced Plasma Spectroscopy and Applications
389 (LIBS 2006), doi:10.1016/j.sab.2007.10.041.
- 390 12. Anabitarte, F.; Mirapeix, J.; Portilla, O.M.C.; Lopez-Higuera, J.M.; Cobo, A. Sensor for the Detection of
391 Protective Coating Traces on Boron Steel With Aluminium–Silicon Covering by Means of Laser-Induced
392 Breakdown Spectroscopy and Support Vector Machines. *IEEE Sensors Journal* **2012**, *12*, 64–70.
393 doi:10.1109/JSEN.2011.2121902.
- 394 13. Bukin, O.; Proshenko, D.; Chekhlenok, A.; Golik, S.; Bukin, I.; Mayor, A.; Yurchik, V. Laser
395 Spectroscopic Sensors for the Development of Anthropomorphic Robot Sensitivity. *Sensors* **2018**, *18*.
396 doi:10.3390/s18061680.
- 397 14. Moncayo, S.; Manzoor, S.; Navarro-Villoslada, F.; Caceres, J.O. Evaluation of supervised chemometric
398 methods for sample classification by Laser Induced Breakdown Spectroscopy. *Chemometrics and Intelligent*
399 *Laboratory Systems* **2015**, *146*, 354–364. doi:10.1016/j.chemolab.2015.06.004.
- 400 15. Zhang, T.; Xia, D.; Tang, H.; Yang, X.; Li, H. Classification of steel samples by laser-induced breakdown
401 spectroscopy and random forest. *Chemometrics and Intelligent Laboratory Systems* **2016**, *157*, 196–201.
402 doi:10.1016/j.chemolab.2016.07.001.
- 403 16. Wang, N.; Wang, X.; Chen, P.; Jia, Z.; Wang, L.; Huang, R.; Lv, Q. Metal Contamination Distribution
404 Detection in High-Voltage Transmission Line Insulators by Laser-induced Breakdown Spectroscopy (LIBS).
405 *Sensors* **2018**, *18*. doi:10.3390/s18082623.

- 406 17. Zhang, C.; Shen, T.; Liu, F.; He, Y. Identification of Coffee Varieties Using Laser-Induced Breakdown
407 Spectroscopy and Chemometrics. *Sensors* **2018**, *18*. doi:10.3390/s18010095.
- 408 18. Tognoni, E.; Palleschi, V.; Corsi, M.; Cristoforetti, G.; Omenetto, N.; et al., I.G., Laser-induced breakdown
409 spectroscopy (LIBS) fundamentals and applications; New York, USA: Cambridge University Press, 2006;
410 chapter From sample to signal in laser-induced breakdown spectroscopy: a complex route to quantitative
411 analysis, p. 122.
- 412 19. Tadeusiewicz, R. Place and Role of Intelligent Systems in Computer Science. *Computer Methods in Materials
413 Science* **2010**, *10*, 193–206.
- 414 20. Pławiak, P.; Sośnicki, T.; Niedźwiecki, M.; Tabor, Z.; Rzecki, K. Hand Body Language Gesture Recognition
415 Based on Signals From Specialized Glove and Machine Learning Algorithms. *IEEE Transactions on Industrial
416 Informatics* **2016**, *12*, 1104–1113. doi:10.1109/TII.2016.2550528.
- 417 21. Pławiak, P. An estimation of the state of consumption of a positive displacement pump based
418 on dynamic pressure or vibrations using neural networks. *Neurocomputing* **2014**, *144*, 471 – 483.
419 doi:https://doi.org/10.1016/j.neucom.2014.04.026.
- 420 22. Pławiak, P.; Maziarz, W. Classification of tea specimens using novel hybrid artificial intelligence methods.
421 *Sensors and Actuators B: Chemical* **2014**, *192*, 117 – 125. doi:10.1016/j.snb.2013.10.065.
- 422 23. Rzecki, K.; Pławiak, P.; Niedźwiecki, M.; Sośnicki, T.; Leśkow, J.; Ciesielski, M. Person recognition based
423 on touch screen gestures using computational intelligence methods. *Information Sciences* **2017**, *415–416*, 70 –
424 84. doi:10.1016/j.ins.2017.05.041.
- 425 24. Pławiak, P.; Rzecki, K. Approximation of Phenol Concentration Using Computational Intelligence
426 Methods Based on Signals From the Metal-Oxide Sensor Array. *Sensors Journal, IEEE* **2015**, *15*, 1770–1783.
427 doi:10.1109/JSEN.2014.2366432.
- 428 25. Pławiak, P. Novel methodology of cardiac health recognition based on ECG signals
429 and evolutionary-neural system. *Expert Systems with Applications* **2018**, *92*, 334 – 349.
430 doi:https://doi.org/10.1016/j.eswa.2017.09.022.
- 431 26. Pławiak, P. Novel genetic ensembles of classifiers applied to myocardium dysfunction
432 recognition based on ECG signals. *Swarm and Evolutionary Computation* **2018**, *39*, 192 – 208.
433 doi:https://doi.org/10.1016/j.swevo.2017.10.002.
- 434 27. Abdar, M.; Zomorodi-Moghadam, M.; Das, R.; Ting, I.H. Performance analysis of classification
435 algorithms on early detection of liver disease. *Expert Systems with Applications* **2017**, *67*, 239 – 251.
436 doi:https://doi.org/10.1016/j.eswa.2016.08.065.
- 437 28. Trejos, T.; Flores, A.; Almirall, J.R. Micro-spectrochemical analysis of document paper and gel inks by
438 laser ablation inductively coupled plasma mass spectrometry and laser induced breakdown spectroscopy.
439 *Spectrochimica Acta Part B: Atomic Spectroscopy* **2010**, *65*, 884 – 895. doi:10.1016/j.sab.2010.08.004.
- 440 29. Kula, A.; Wietecha-Posłuszny, R.; Pasionek, K.; Król, M.; Woźniakiewicz, M.; Kościelniak, P. Application of
441 laser induced breakdown spectroscopy to examination of writing inks for forensic purposes. *Science &
442 Justice* **2014**, *54*, 118 – 125. doi:10.1016/j.scijus.2013.09.008.
- 443 30. Elsherbiny, N.; Nassef, O.A. Wavelength dependence of laser induced breakdown spectroscopy (LIBS) on
444 questioned document investigation. *Science & Justice* **2015**, *55*, 254 – 263. doi:10.1016/j.scijus.2015.02.002.
- 445 31. Lennard, C.; El-Deftar, M.M.; Robertson, J. Forensic application of laser-induced breakdown spectroscopy
446 for the discrimination of questioned documents. *Forensic Science International* **2015**, *254*, 68 – 79.
447 doi:10.1016/j.forsciint.2015.07.003.
- 448 32. Hoehse, M.; Paul, A.; Gornushkin, I.; Panne, U. Multivariate classification of pigments and inks using
449 combined Raman spectroscopy and LIBS. *Analytical and Bioanalytical Chemistry* **2011**, *402*, 1443–1450.
450 doi:10.1007/s00216-011-5287-6.
- 451 33. Metzinger, A.; Rajkó, R.; Galbács, G. Discrimination of paper and print types based on their laser
452 induced breakdown spectra. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2014**, *94–95*, 48 – 57.
453 doi:http://dx.doi.org/10.1016/j.sab.2014.03.006.
- 454 34. Team of Science and Industrial Intelligent Applications; subsection “Research” in:
455 35. Zieliński, T.P. *Digital signal processing: from theory to applications*; WK Ł, 2005.
- 456 36. Murphy, K.P. *Machine learning: a probabilistic perspective*, 1st ed.; The MIT Press: Cambridge, MA, 2012.
- 457 37. Specht, D.F. A general regression neural network. *IEEE Transactions on Neural Networks* **1991**, *2*, 568–576.
458 doi:10.1109/72.97934.

- 459 38. Specht, D.F. Probabilistic neural networks. *Neural Networks* **1990**, *3*, 109–118.
460 doi:doi.org/10.1016/0893-6080(90)90049-Q.
- 461 39. Hinton, G.E. Connectionist Learning Procedures. *Artif. Intell.* **1989**, *40*, 185–234.
462 doi:10.1016/0004-3702(89)90049-0.
- 463 40. Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995**, *20*, 273–297.
464 doi:10.1007/BF00994018.
- 465 41. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*;
466 Springer Publishing Company, Incorporated, 2014.
- 467 42. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American*
468 *Statistician* **1992**, *46*, 175–185. doi:10.1080/00031305.1992.10475879.
- 469 43. Sugeno, M. *Industrial applications of fuzzy control*; Elsevier Science Pub. Co., 1985.
- 470 44. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine*
471 *Learning)*; The MIT Press, 2005.
- 472 45. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.;
473 Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E.
474 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
- 475 46. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognition Letters* **2006**, *27*, 861–874.
476 doi:doi.org/10.1016/j.patrec.2005.10.010.
- 477 47. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks.
478 *Information Processing & Management* **2009**, *45*, 427 – 437. doi:10.1016/j.ipm.2009.03.002.
- 479 48. Trejos, T.; Corzo, R.; Subedi, K.; Almirall, J. Characterization of toners and inkjets by
480 laser ablation spectrochemical methods and Scanning Electron Microscopy-Energy Dispersive
481 X-ray Spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2014**, *92*, 9 – 22.
482 doi:https://doi.org/10.1016/j.sab.2013.11.004.
- 483 49. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *CoRR* **2014**, *abs/1412.6980*, [1412.6980].
- 484 50. Chui, C.K. *An Introduction to Wavelets*; Academic Press Professional, Inc.: San Diego, CA, USA, 1992.

485 **Sample Availability:** LIBS spectra used in this study are available on the webpage: <http://libs.iti.pk.edu.pl/>.