



Communication

Incremental Learning in Modelling Process Analysis Technology (PAT)—An Important Tool in the Measuring and Control Circuit on the Way to the Smart Factory

Shivani Choudhary ^{1,*†}, Deborah Herdt ^{1,*†} , Erik Spoor ¹, José Fernando García Molina ², Marcel Nachtmann ¹ and Matthias Rädle ¹ 

- ¹ Center for Mass Spectrometry and Optical Spectroscopy, Mannheim University of Applied Sciences, Paul-Wittsack-Straße 10, 68163 Mannheim, Germany; e.spoor@hs-mannheim.de (E.S.); m.nachtmann@hs-mannheim.de (M.N.); m.raedle@hs-mannheim.de (M.R.)
- ² Institute of Process Control and Innovative Energy Conversion, Mannheim University of Applied Sciences, 68163 Mannheim, Germany; Jose.Fernando.Garcia.Molina@continental-corporation.com
- * Correspondence: s.choudhary@hs-mannheim.de (S.C.); d.herdt@hs-mannheim.de (D.H.)
- † These authors contributed equally to this work.



Citation: Choudhary, S.; Herdt, D.; Spoor, E.; García Molina, J.F.; Nachtmann, M.; Rädle, M. Incremental Learning in Modelling Process Analysis Technology (PAT)—An Important Tool in the Measuring and Control Circuit on the Way to the Smart Factory. *Sensors* **2021**, *21*, 3144. <https://doi.org/10.3390/s21093144>

Academic Editors: Fabian Khateb, Bernhard Wilhelm Roth, Krzysztof M. Abramski and Guillermo Villanueva

Received: 2 March 2021
Accepted: 28 April 2021
Published: 1 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: To meet the demands of the chemical and pharmaceutical process industry for a combination of high measurement accuracy, product selectivity, and low cost of ownership, the existing measurement and evaluation methods have to be further developed. This paper demonstrates the attempt to combine future Raman photometers with promising evaluation methods. As part of the investigations presented here, a new and easy-to-use evaluation method based on a self-learning algorithm is presented. This method can be applied to various measurement methods and is carried out here using an example of a Raman spectrometer system and an alcohol-water mixture as demonstration fluid. The spectra's chosen bands can be later transformed to low priced and even more robust Raman photometers. The evaluation method gives more precise results than the evaluation through classical methods like one primarily used in the software package Unscrambler. This technique increases the accuracy of detection and proves the concept of Raman process monitoring for determining concentrations. In the example of alcohol/water, the computation time is less, and it can be applied to continuous column monitoring.

Keywords: SVM; incremental learning; Raman spectroscopy; process technology

1. Introduction

Process analysis technology (PAT) is established in many chemical industry plants. It enables the production of the required technical quality in compliance with safety standards. This is made possible with the best possible use of raw materials, systems, and energy. The chemical industry is the most energy-intensive manufacturing industry in Germany, with consumption of 1137.3 petajoules in 2018 [1]. This corresponds to a work or energy of 36.06 gigawatt years. From an economic perspective alone, resource efficiency is of great importance. From 1990 to 2018, the chemical–pharmaceutical industry was able to increase its production by 76%, reduce energy consumption by 17%, and reduce greenhouse gas emissions by 51% [2]. It is of the utmost importance to use sensors that significantly increase process understanding and allow more profound insight into the process. The profit is immense, especially for existing systems. Direct in-line measurement of the current composition for regulation is currently rarely used due to the high costs. Therefore, the typical sensors in-process application focus on process control variables—such as pressure and temperature—instead of in-line-product variables. However, measuring the composition of substances is particularly valuable to optimize processes economically and energetically. Optical processes are growing slowly because knowledge is obtained in-line and no extractions of the process have to be made, but accuracy is too low and the

price too high. These systems must be flexible and quick to use to understand the process further. Permanent measuring points must be characterized by high measuring accuracy and reliability with low installation and operating costs. To guarantee process stability and, therefore, a high performance, the process has to be monitored. With high measurement density, digital output of the data, and numerous individual and real-time measurements, the system's state becomes more predictable. Additionally, simultaneous measurements of several indicators of the process stabilize the predicted status, and consequently, rapid intervention, if necessary, is ensured.

A possible and suitable method for those mentioned challenges is Raman technology with high selectivity and detection sensitivity. The integration into the process can be realized with a flange connector mounted on an inspection glass in a pipe. The excitation wavelength from the laser and the scattered light from the sample can pass through the window [3,4]. With a calibration model, the resulting Raman shift intensities can be converted into the substance's concentration. This continuous measurement variable leads to sustainable and safe controlled operations. From an economic point of view, the plant's productivity can be increased or decreased as required. Therefore, this method can be adjusted according to demand, changing energy and raw material prices, leading to enhanced profit. Production efficiency is an increasingly differentiating characteristic for companies. This characteristic is per the sustainability requirements. Worldwide, there is a demand for conserving resources, reducing global warming gases, and, therefore, sustainable energy usage.

In the present work, an incremental learning evaluation model using a Support Vector Machine (SVM) model for Raman spectroscopic data is presented, which ensures user-friendly recalibration during operation. Its performance is determined experimentally and compared with conventional modelling techniques. SVM are robust classifiers, but large datasets lead to long computation times, high memory requirements, and increased complexity of the model. To solve this issue, SVM ensembles, where each SVM sees only a fraction of the data, are a viable solution [5]. Standard methods are used to achieve a high accuracy classifier by computing the best hyperparameters for the SVM model like tenfold cross-validation and grid search [6,7].

Generally, the studies using SVM learning from new data involve discarding the existing classifier, integrating the new data to the old set, and training a new classifier from scratch. The studies do not learn incrementally with the addition of new data, and they result in unlearning of data [5,8]. This means that the system cannot learn new information without forgetting previously learned classifiers. Such a problem is solved with the help of an incremental learning algorithm, defined as one that meets the following criteria [9,10]:

1. Can learn additional information from new data
2. Does not require access to the original data used to train the existing classifier
3. Preserves previously acquired knowledge

Additionally, the proposed incremental learning systems [11–14] suffer from high computation time and complexity. This framework's main contribution is the implementation and evaluation of an incremental learning algorithm based on Garcia et al. [15], which uses parallel computing and helps reduce the computation time with reduced complexity as opposed to previously used learning algorithms [16]. The developed method's relevance in the industrial environment is represented and discussed from a technical perspective. In the example process of rectification, a feed stream of ethanol and water is thermally separated, and ethanol is removed overhead as a distillate. Water leaves the column via the sump stream.

2. Materials and Methods

2.1. Experimental Setup Raman-Spectroscopy

Raman Spectroscopy was performed with a tec5 MultiSpec[®]Raman system (tec5, Steinbach, Germany), equipped with a coaxial designed RamanProbe II (InPhotonics, Norwood, Massachusetts, USA; fibre configuration: 105 µm excitation fibre; 600 µm collection

fibre/working distance 7.5 mm). Raman scattering was excited by a 500 mW temperature-stabilized semiconductor laser (Raman Boxx™, PD-LD500) source at 785 nm. The Raman spectra were collected with a 1 cm^{-1} resolution ranging from $300\text{--}3215\text{ cm}^{-1}$ using MultiSpec Pro II Raman process software (v1.4.1189.1826, tec5, Steinbach, Germany). The probe was clamped with a laboratory stand in the 50 mL borosilicate beaker (from Schott, Mainz, Germany). Figure 1 shows the setup schematically. Each solution (1–10) was measured ten times with an integration time of 10 s.

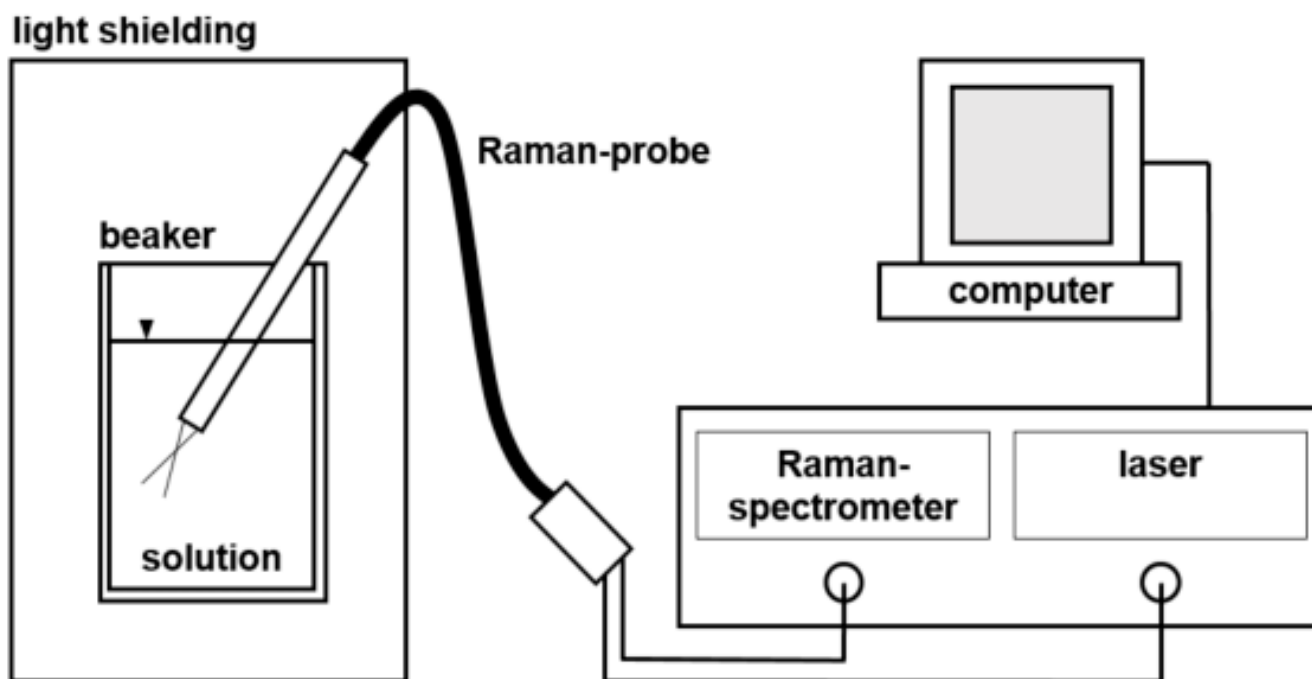


Figure 1. Experimental setup of Raman measurement.

The dilution series was prepared with 96.2 Vol% Ethanol from VWR (Darmstadt, Germany, CAS No. 64-17-5) and distilled water. In total, ten solutions were prepared in 50 mL volumetric flasks (from Schott, Mainz, Germany). Initially, ethanol was transferred, with an Eppendorf Reference pipette (100-1000 μ L; Eppendorf AG, Hamburg, Germany), into the volumetric flask, and consequently, the remaining volume was filled up with distilled water according to Table 1. Solutions were then transferred into a 50 mL beaker for subsequent measurement.

Table 1. Dedicated samples numbers with calculated ethanol concentration.

Sample Number	Volume H ₂ O/mL	Volume EtOH/mL	Calculated EtOH Concentration/Vol%
0	50.00000000	0.00000000	0.00000000
1	49.50000000	0.50000000	0.96200000
2	49.75000000	0.25000000	0.48100000
3	49.87500000	0.12500000	0.24050000
4	49.93750000	0.06250000	0.12025000
5	49.96875000	0.03125000	0.06012500
6	49.98437500	0.01562500	0.03006250
7	49.99218750	0.00781250	0.01503125
8	49.99609375	0.00390625	0.007515625
9	49.998046875	0.001953125	0.003757813

2.2. Algorithm

In this work, the developed data evaluation method's advantages were presented, which ultimately resulted in lowering the detection limit and increased robustness against outliers or poorly representative data. As a method, a learning algorithm was developed that has particular advantages concerning the learning speed in the continuous expansion of the database when new data were added. This was achieved by taking samples during monitoring as well as analyzing them in the laboratory. In time, they can be integrated into the model. The model improved continuously with accumulating data, which resulted in increased accuracy and a decreased error rate. The mathematical model used in this paper was adapted from the works of Garcia et al. [15].

Current computer-aided tools have made strides towards accurate and efficient detection of chemical concentrations. However, a common drawback observed in these approaches was the use of models that suffer from unlearning [5,8]. In these models, the previously learned knowledge was discarded, and new models had to be trained from scratch in the learning phase as soon as new data became available. In actual process conditions, in which the trained data were presented with time delay over a period of time, the standard gold data had a lack of stability. This was due to the sample taking influences, inhomogeneous product distribution, and impurities. This gave the reason that new emerging areas in machine learning systems must be investigated.

The advantage of an incremental learning algorithm was that it could learn additional information step by step as soon as new data were available. The ensemble learning algorithm used was implemented based on the work of Robi Polikar et al. [9]. The data set was divided into two classes: The first class corresponded to the spectra with a dilution greater than or equal to a specific dilution limit, and the second class contained the dilutions below this limit.

2.3. Mathematics

Based on the dynamically updated distribution of the training data set, the ensemble classifier was trained so that the samples that were more difficult to classify were given an increased probability of increasing their chances of being selected in the following training data set. The algorithm used the database $DB_k, k = 1, \dots, K$ where K represents the number of available measurement series, in this work ten. The random samples of all measurement series were first permuted arbitrarily and divided into $K = 10$ stacks of equal size.

The inevitable case wording of SVM was used as the primary classifier, which is referred to below as 1C-SVM. The goal of training with the 1C-SVM was to establish an optimal hypothesis h with which the two classes were separated. The distance between the dividing line or hypothesis and the training data (called support vectors) of each class was maximized, and thus the model was protected against incorrect specifications and increased the robustness of the forecast. The goal was to determine an optimal hyperplane $f(x) = \varphi(x_i)^T w + w_0$ in the feature space. The following formula was used for this:

$$\min_{w, w_0, \varepsilon} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \varepsilon_i \quad (1)$$

$$\begin{aligned} &\text{Subject to } \varepsilon_i \geq 0 \\ &y_i \left(\varphi(x_i)^T w + w_0 \right) \geq 1 - \varepsilon_i \quad \forall i \end{aligned}$$

In it forms $\varphi(x_i)$ maps x_i in a higher-dimensional space. w is the weight vector. C is a regulated hyperparameter that added a penalty to the target function in the event of overfitting. ε is a slip variable to weaken the limits and is called an error range or misclassification error. The Equation (1) can be determined using a quadratic approach with a second kind Lagrangian function [17]. By using this function, the kernel trick $K(x_i, x_j)$ can be used. With this it was possible to transfer the data to a higher dimension

so that non-linear dependencies between different training data can be considered. The mapping in a higher dimensionality enabled the detection of similarities between data characteristics. The Gaussian radial basis function was used as the kernel, and accordingly, the following Equation (2) was obtained:

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right), \gamma = \frac{1}{2\sigma^2} \quad (2)$$

σ is the variance. γ is a hyperparameter that smoothed the kernel function, and this means a stronger or weaker relationship between the samples can be found depending on the γ values. To estimate the optimal hyperparameters γ and C for 1C-SVM, Nelder-Mead's heuristic method was used [18,19]. The selected criterion to find the optimal classifier was the area under the Receiver Operating Characteristic curve (AUC-ROC), which is increasingly used in machine learning and evaluating more significant amounts of data.

The inputs for the ensemble algorithm were:

1. Training data $S_k = \{(x_i, y_i) | i = 1, \dots, N_k\}$. The data set consisted of N_k training data points with $x_i \in R^d$, where d represents the number of dimensions and $y_i \in \{-1, 1\}$, the associated class. The N_k data points were randomly selected from the k^{th} database (DB_k).
2. A primary classifier to generate hypothesis h . The classifier required that at least 50 percent of the training data record was classified correctly.
3. An integer T_k that specified the number of iteration steps $t = 1, 2, \dots, T_k$ for each data set, with $t \in N$. The prediction error could be reduced sufficiently with T_k .

The ensemble algorithm started with the initialization of a series of weights α_t for the training data set S_k and a distribution D_t obtained from α_t [1]. According to D_t , S_k was divided into two subsets, TR_t for training and TE_t for validating during the t^{th} iteration of the algorithm. D_t was initially defined uniformly without deductive information. At each iteration, the weights adjusted at iteration $t - 1$ were divided by the sum of all $W_t - 1$ to ensure a legitimate distribution, and a new D_t was computed. Training and test subsets were drawn randomly according to D_t . A hypothesis h_t was obtained as the t^{th} classifier, whose error ε_t (3) was computed on the entire data set S_k with:

$$\varepsilon_t = \frac{\sum_{i:h_t(x_i)} D_t \cdot |y_i - h_t(x_i)|}{\sum_{i:h_t(x_i)} D_t} \quad (3)$$

ε_t was required to be less than 0.5 to ensure a reasonable performance of h_t . If the condition was satisfied, h_t was accepted, and the error was normalized to calculate the normalization error β_t (4):

$$\beta_t = \frac{\varepsilon_t}{(1 - \varepsilon_t)}, 0 < \beta_t < 1 \quad (4)$$

The current h_t was discarded if the condition was not satisfied, and a new training subset was selected. All hypotheses generated so far were then combined using the weighted majority voting to obtain a composite one H_t (5), which allowed efficient incremental learning capability when new classes were introduced. The hypothesis with good performance was awarded a higher voting weight [20].

$$H_t = \operatorname{argmax}_{y \in Y} \sum_{t:h_t(x)=y} \log\left(\frac{1}{\beta_t}\right) \quad (5)$$

The error of H_t was computed with (6) and must have also been less than 0.5 to ensure a reasonable performance of H_t ; otherwise, the algorithm discarded that one and returned to select a new TR_t .

$$E_t = \frac{\sum_{i:H_t(x_i)} D_t \cdot |y_i - H_t(x_i)|}{\sum_{i:H_t(x_i)} D_t} \quad (6)$$

Then B_t was computed with (7).

$$B_t = \frac{E_t}{(1 - E_t)}, \quad 0 < B_t < 1 \quad (7)$$

The rule of Equation (8) was used to reduce the weights of those data points that were correctly classified by the composite hypothesis H_t . Furthermore, this lowered the probability of being selected in the following training subset.

$$\alpha_{t+1}(i) = \alpha_t(i) \cdot \begin{cases} B_t, & \text{if } H_t(x_i) = y_i \\ 1, & \text{otherwise} \end{cases} \quad (8)$$

The hypothesis H_F for the training subset and the subset of features could be obtained by combining all hypotheses that had been generated so far using the weighted majority voting rule (see Figure 2).

$$H_F(x) = \operatorname{argmax}_{y \in Y} \sum_{k=1}^K \sum_{t: H_t(x)=y} \log\left(\frac{1}{B_t}\right) \quad (9)$$

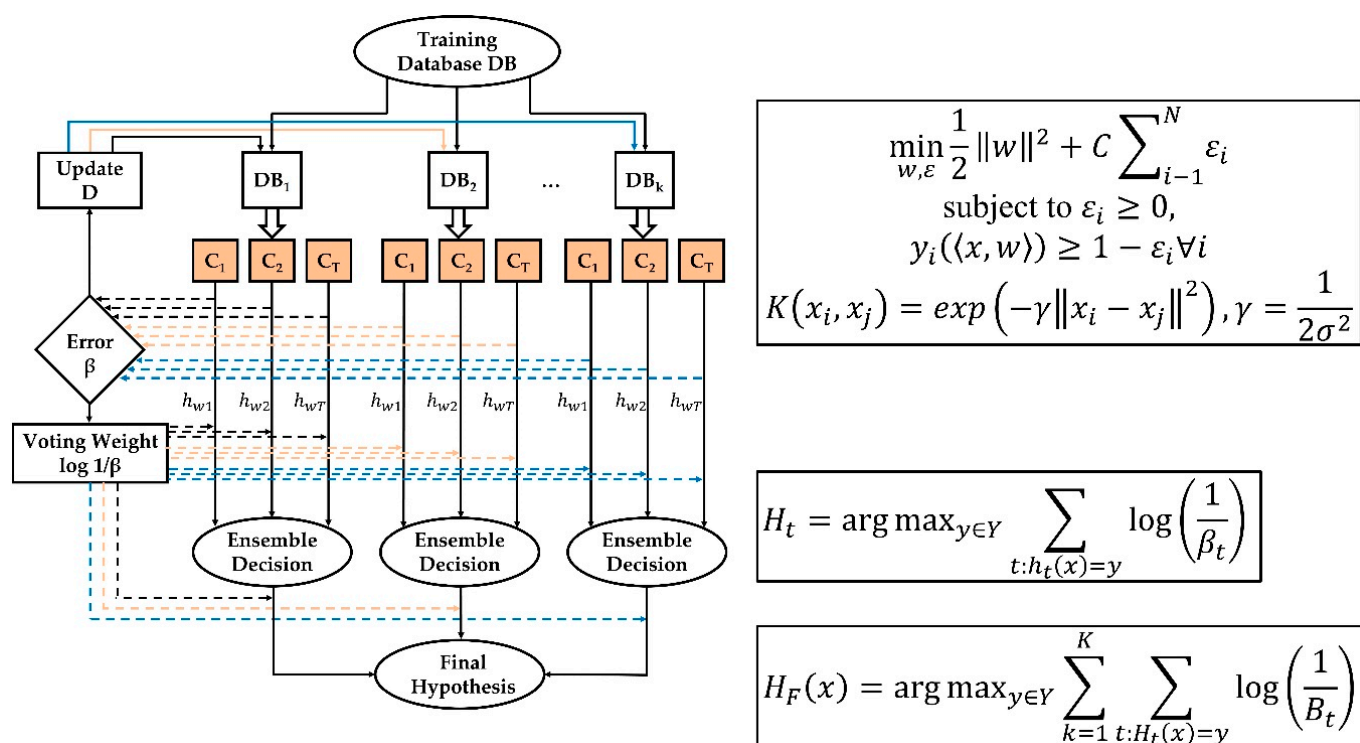


Figure 2. Schematic representation of the incremental learning ensemble classifier (altered from [20]).

2.4. Evaluation

The water spectrum was subtracted from each sample's spectrum. Then, 454 main features from intervals of the spectrum were used to evaluate the data. The intervals contained the spectrum's descriptive peaks, which were between the Raman shifts 850–910, 1010–1130, 1410–1510, and 2840–3010 cm^{-1} . These characteristics corresponded to the value of the derivative and the integral in each interval.

The following methodology was implemented to obtain an objective evaluation. The data set was divided into three subsets: training, validation, and test. The validation subset aimed to make a fair estimation of performance, independency of the test data, and to pick

optimal parameters, which in turn provided a more generalized solution. We implemented a nested cross-validation (CV) method for unbiased estimation of prediction error in an independent test data. Leave-one-out (LOO) was used to evaluate the classifier, i.e., nine experiments were used for training and validation and one experiment for independent testing. To estimate the classifier's unknown tuning parameters, tenfold CV was used with data points from the nine experiments selected for training and validation. The data points were randomly permuted and divided in ten parts, one part for the validation and nine for training. Statistical measures such as the sensitivity true positive results (TPR) and precision were also computed (see Equation (10)). A thresholding procedure was used to design a binary classification. The threshold T_h was selected as the concentration level such that a particular concentration was higher than the threshold concentration if the probability $p\left(\frac{y_i}{x_i}\right) \geq T_h$ else it was classified as lower concentration than required, and a ground truth table based on this idea was generated for every concentration. The use of TPR and Positive Predictive Value (PPV) provided additional information about the classifier's performance and could be used to compare results obtained with other methods. Additional parameters like the PPV and TPR were also calculated as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

TP and TN denote the number of true positive and true negative data points, respectively. FP and FN denote the number of false-positive and false-negative data points, respectively.

The SVM model in The Unscrambler X version 10.4 (Camo Software, Oslo, Norway) was used to compare the algorithm with a standard program. The ground truth table generated using the MATLAB algorithm was used as input data for binary classification. For each concentration, cross-validation was performed in the same way the MATLAB algorithm does. This resulted in the creation of nine models and predictions per concentration.

3. Results

Figure 3 shows the resulting Raman spectra of ethanol and water in the complete recorded region. The descriptive peaks for ethanol in between 850–910, 1010–1130, 1410–1510, and 2840–3010 cm^{-1} were used in the algorithm. Additionally, two Raman peaks at 435 and 1275 cm^{-1} were also visible in the ethanol spectrum. Water showed typical bands at 500 cm^{-1} (hydrogen bond), 1640 cm^{-1} (OH bending), and 3100–3600 cm^{-1} (OH stretching). Remaining peaks were derived from the glass beaker.

In Figure 4, the spectral data of the dilution series is displayed. For reasons of clarity, the range in between 850–1130 cm^{-1} was enlarged to show the linear concentration dependency more clearly.

The outputs from the predictions obtained through Unscrambler X must be categorised manually into true positive, true negative, false positive, and false negative to calculate the mean value for accuracy, precision, and recall (sensitivity) for each concentration. The results obtained through Unscrambler X are shown in Table 2.

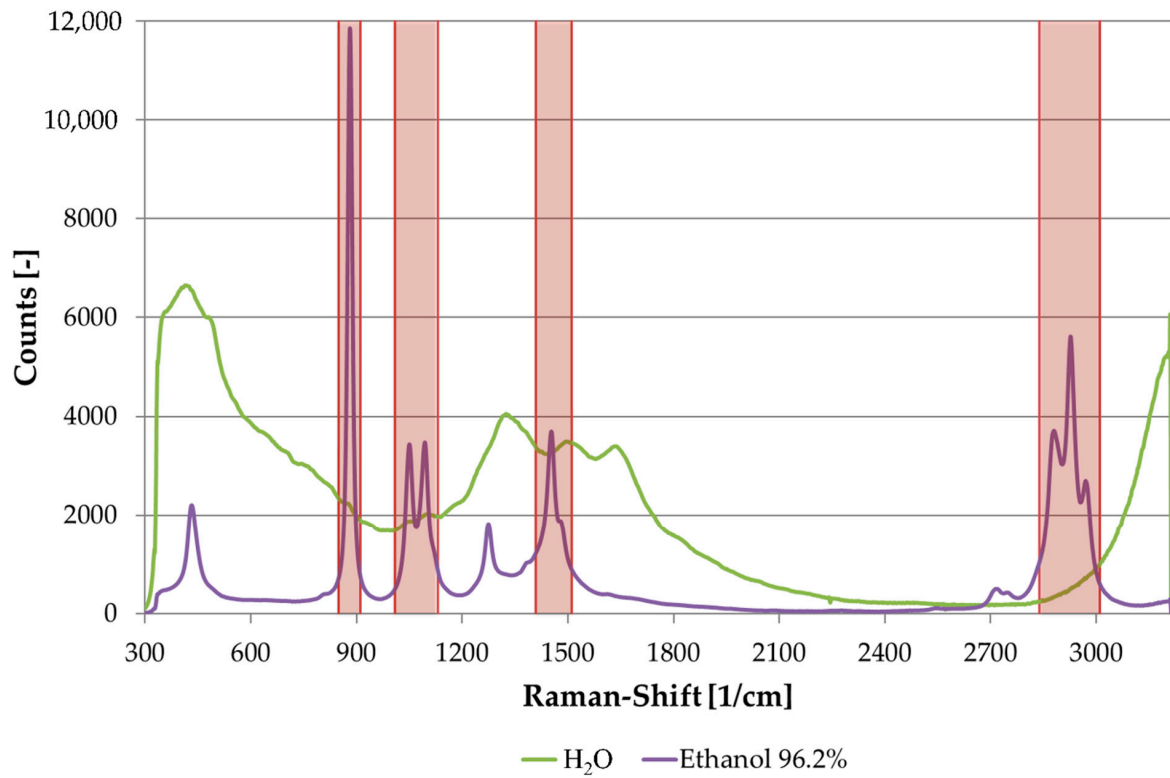


Figure 3. Raman spectra of water and ethanol with evaluated intervals.

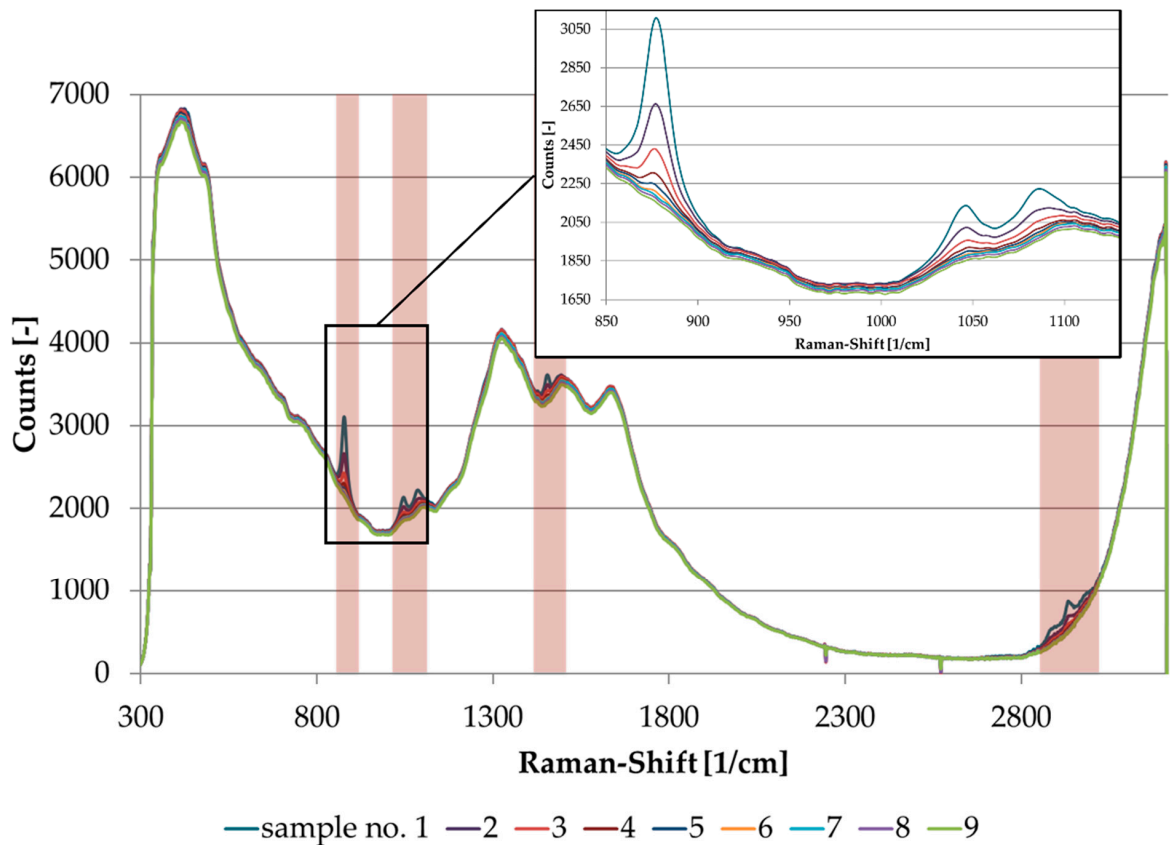


Figure 4. Raman spectra of ethanol dilution series (sample number 1-9).

Table 2. Calculated mean value for accuracy, precision, recall/sensitivity, and computation time for each concentration with the c-value and gamma obtained from the algorithm using Unscrambler X.

Calculated EtOH Concentration (Vol%)	Accuracy Unscrambler (Unscrambler Parameters) (%)	Precision (-)	Recall/Sensitivity (-)
0.9620	100.0	1.00	1.00
0.4810	100.0	1.00	1.00
0.2405	100.0	1.00	1.00
0.1203	98.7	0.99	0.98
0.0601	95.8	0.95	0.97
0.0301	90.9	0.95	0.90
0.0150	83.4	0.90	0.87
0.0075	84.7	0.90	0.92
0.0038	90.7	0.93	0.98

Table 3 displays the results obtained from the MATLAB algorithm. The different parameters for the performance of the classifier were computed using the algorithm. The training time, as well as the classification time, was calculated for each concentration. The time taken for validation and training was also calculated.

Table 3. Calculated mean value for accuracy, precision, recall/sensitivity, and time required for training and validation for each concentration obtained from the MATLAB algorithm.

Calculated EtOH Concentration (Vol%)	Accuracy MATLAB Algorithm (MATLAB Algorithm Parameters) (%)	Precision (-)	Recall/Sensitivity (-)	Training and Validation Time (s)
0.9620	100.0	1.00	1.00	14.5748
0.4810	100.0	1.00	1.00	17.8897
0.2405	99.9	0.99	1.00	43.3492
0.1203	99.9	0.98	0.98	40.5498
0.0601	99.0	0.97	0.97	148.2467
0.0301	96.2	0.96	0.95	225.5823
0.0150	91.9	0.97	0.93	140.3069
0.0075	89.1	0.98	0.93	92.7786
0.0038	83.7	0.99	0.94	48.4829

As can be observed from the results obtained, when the concentration is higher, both the methods' performances were more or less identical. However, as the concentration decreases, the MATLAB algorithm showed better accuracy, precision, and sensitivity. For concentrations from 0.962 to 0.12025, the performance with respect to precision and sensitivity of both methods was almost the same, but the MATLAB algorithm's accuracy was still higher. As we go to lower concentrations from 0.060125 to 0.007515625, there was a significant difference between the accuracy, precision, and sensitivity from the MATLAB algorithm and Unscrambler X. This shows that the MATLAB algorithm gave better results in terms of accuracy, precision, and sensitivity.

In Figure 5, the accuracy from The Unscrambler X and the accuracy from the MATLAB algorithm are plotted over concentration. For concentrations from sample number 1–3, the Unscrambler X models with the MATLAB parameters showed a significantly lower accuracy, while the MATLAB algorithm and the Unscrambler X model with the new parameters showed an accuracy of 100%. This implied that the Unscrambler X and MATLAB work with different parameters that cannot be taken from each other to create the same results. For calculated ethanol concentrations lower than 0.481 Vol% (sample number 3), the MATLAB algorithm generated higher accuracy and displayed slighter decrease in accuracy over the samples. At a calculated ethanol concentration of 0.0300625 Vol% (sample number 7), the

MATLAB algorithm had the highest difference in accuracy compared to the Unscrambler model with approximately 8%.

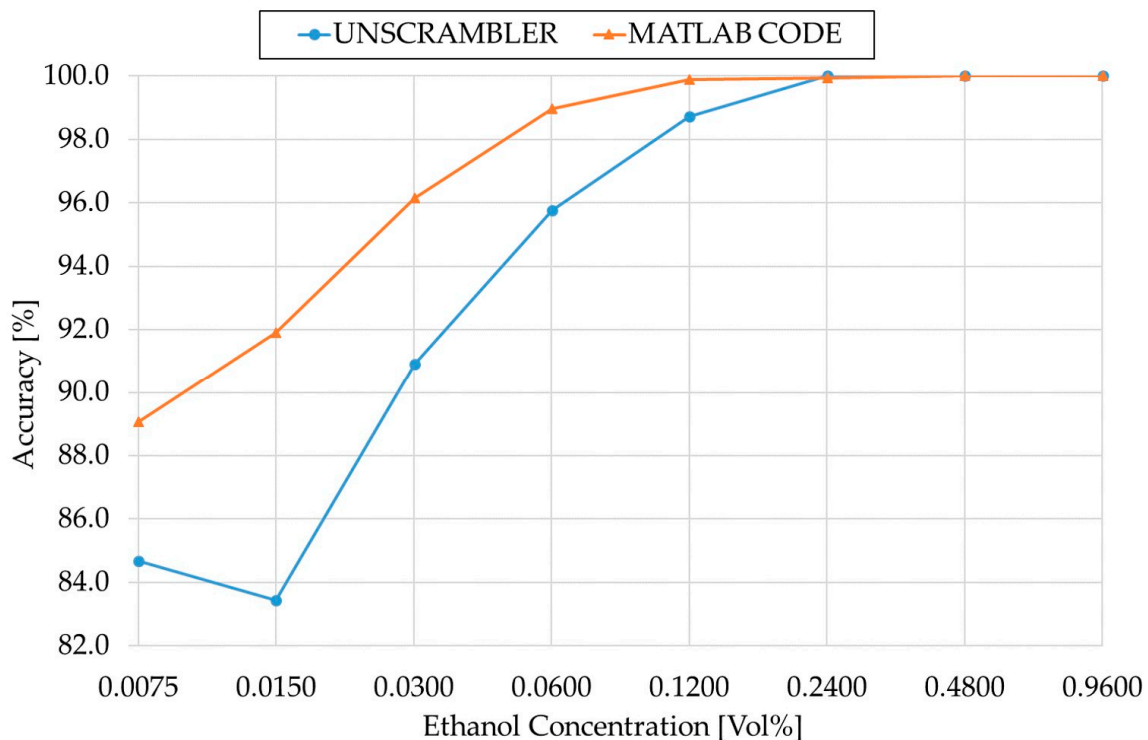


Figure 5. Comparison of the accuracy from The Unscrambler X and the MATLAB code.

4. Discussion

To solve the problem of unlearning in models used in PAT processes, pattern recognition systems that can be better adapted, scalable, and given the ability to learn different data distributions dynamically should be used, but are not available. The method for incremental learning algorithm that uses an ensemble algorithm for classification in a rectification process of ethanol is a possible solution for unlearning. It can learn additional information as soon as new data are available without unlearning the previously acquired knowledge and prevents the unlearning of old Raman spectroscopic data. The system is continuously updated with the current data and significantly increases robustness and accuracy without the loss of any data.

The main advantage of the proposed method over the standard methods available, like Unscrambler X, was that the computation time was less, accuracy was higher, and manual intervention was not involved. The disadvantage in the SVM model created using Unscrambler X was that the data had to be prepared manually, and for each concentration, a model had to be created. The proposed method works without manual intervention saving the users a lot of time and effort. Additionally, a Grid Search had to be performed for each concentration in Unscrambler X. This resulted in a total of 81 models that had to be evaluated manually. Hence, the total time to calculate the accuracy for all concentrations depended on the operator's efficiency and speed.

The algorithm of the developed new model's self-learning follow-up enabled the computing time (batch processing, parallel execution, and distributed data) to be reduced. Additionally, the accuracy and the detection sensitivity of the measuring devices that are used can be increased. In this paper, the spectrum range peaks were selected and used, but the algorithm's application can be extended to any number of peaks and any range in the spectrum that suits the application. The algorithm can be adapted for different data and

applications where continuous monitoring is required with the addition of new data. The algorithm performs better than standard methods until the concentration is above 0.0075.

Further investigations need to be performed using different ranges of spectroscopic data and with different chemical processes. The method presented further increases the robustness against outliers or poorly representative data, which do occur in the measured values obtained during operation determined by the system's operating personnel. The fast availability of measurement data increases the system flexibility and thus the constant optimization to changing demands. To the best of our knowledge, the present work is the first to apply the incremental learning SVM model in Raman spectroscopic data used in various chemical processes where continuous monitoring is required.

Author Contributions: Conceptualization: D.H. and S.C.; methodology: D.H. and S.C.; software: J.F.G.M., S.C., and E.S.; validation: D.H. and S.C.; formal analysis: D.H., S.C. and E.S.; investigation: D.H. and M.N.; writing—original draft preparation: D.H. and S.C.; writing—review and editing: D.H. and S.C.; visualization: D.H. and S.C.; supervision: M.R.; funding acquisition: M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Bundesministerium für Wirtschaft und Energie (BMWi) and the Arbeitsgemeinschaft industrieller Forschungsvereinigungen (AiF, ZF4013934RE7).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from José Fernando García Molina's thesis and is available from the authors with the permission of García Molina.

Acknowledgments: The authors like to thank Frank Braun for his technical support.

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

References

1. Bretkopf, A. Energieverbrauch in Deutschland Nach Wirtschaftszweig 2018: Energieverbrauch Des Verarbeitenden Gewerbes in Deutschland Nach Ausgewählten Sektoren im Jahr 2018. Available online: <https://de.statista.com/statistik/daten/studie/432596/umfrage/energieverbrauch-im-verarbeitenden-gewerbe-in-deutschland-nach-sektor/> (accessed on 31 March 2020).
2. Hohmann, M. Entwicklung von Produktion, Energieverbrauch und Treibhausgasemissionen der Chemisch-Pharmazeutischen Industrie in Deutschland im Zeitraum 1990 bis 2018. Available online: <https://de.statista.com/statistik/daten/studie/186496/umfrage/entwicklung-der-chemischen-industrie-in-deutschland/> (accessed on 31 March 2020).
3. Schalk, R.; Braun, F.; Frank, R.; Rädle, M.; Gretz, N.; Methner, F.-J.; Beuermann, T. Non-contact Raman spectroscopy for in-line monitoring of glucose and ethanol during yeast fermentations. *Bioprocess Biosyst. Eng.* **2017**, *40*, 1519–1527. [CrossRef] [PubMed]
4. Braun, F.; Schalk, R.; Brunner, J.; Eckhardt, H.S.; Theuer, M.; Veith, U.; Hennig, S.; Ferstl, W.; Methner, F.-J.; Beuermann, T.; et al. Nicht-invasive Prozesssonde zur Inline-Ramananalyse durch optische Schaugläser. *tm—Tech. Messen* **2016**, *83*. [CrossRef]
5. Ai, H.; Wu, X.; Zhang, L.; Qi, M.; Zhao, Y.; Zhao, Q.; Zhao, J.; Liu, H. QSAR modelling study of the bioconcentration factor and toxicity of organic compounds to aquatic organisms using machine learning and ensemble methods. *Ecotoxicol. Environ. Saf.* **2019**, *179*, 71–78. [CrossRef] [PubMed]
6. Zaidi, S. Novel application of support vector machines to model the two phase boiling heat transfer coefficient in a vertical tube thermosiphon reboiler. *Chem. Eng. Res. Des.* **2015**, *98*, 44–58. [CrossRef]
7. Li, M.; Wei, D.; Liu, T.; Liu, Y.; Yan, L.; Wei, Q.; Du, B.; Xu, W. EDTA functionalized magnetic biochar for Pb(II) removal: Adsorption performance, mechanism and SVM model prediction. *Sep. Purif. Technol.* **2019**, *227*, 115696. [CrossRef]
8. French, R. Catastrophic forgetting in connectionist networks. *Trends Cognit. Sci.* **1999**, *3*, 128–135. [CrossRef]
9. Polikar, R.; Upda, L.; Upda, S.S.; Honavar, V. Learn++: An incremental learning algorithm for supervised neural networks. *IEEE Trans. Syst. Man Cybern. C* **2001**, *31*, 497–508. [CrossRef]
10. Erdem, Z.; Polikar, R.; Gurgen, F.; Yumusak, N. Ensemble of SVMs for Incremental Learning. In *Multiple Classifier Systems, Proceedings of the 6th International Workshop, MCS 2005, Seaside, CA, USA, 13–15 June 2005*; Oza, N.C., Ed.; Springer: Berlin, Germany, 2005; pp. 246–256. ISBN 978-3-540-26306-7.
11. Kho, J.B.; Lee, W.; Choi, H.; Kim, J. An incremental learning method for spoof fingerprint detection. *Expert Syst. Appl.* **2019**, *116*, 52–64. [CrossRef]
12. Li, Y.; Su, B.; Liu, G. An Incremental Learning Algorithm for SVM Based on Combined Reserved Set. *J. Shanghai Jiaotong Univ.* **2016**. [CrossRef]

13. Liu, J.; Pan, T.-J.; Zhang, Z.-Y. Incremental Support Vector Machine Combined with Ultraviolet-Visible Spectroscopy for Rapid Discriminant Analysis of Red Wine. *J. Spectrosc.* **2018**, *2018*, 4230681. [[CrossRef](#)]
14. Ardakani, M.; Escudero, G.; Graells, M.; Espuña, A. Incremental Learning Fault Detection Algorithm Based on Hyperplane-Distance. In *26th European Symposium on Computer Aided Process Engineering: Part A*; Kravanja, Z., Ed.; Elsevier Science: Amsterdam, The Netherlands, 2016; pp. 1105–1110. ISBN 9780444634283.
15. García Molina, J.F.; Zheng, L.; Sertdemir, M.; Dinter, D.J.; Schönberg, S.; Rädle, M. Incremental learning with SVM for multimodal classification of prostatic adenocarcinoma. *PLoS ONE* **2014**, *9*, e93600. [[CrossRef](#)] [[PubMed](#)]
16. Xu, G.; Zhou, H.; Chen, J. CNC internal data based incremental cost-sensitive support vector machine method for tool breakage monitoring in end milling. *Eng. Appl. Artif. Intell.* **2018**, *74*, 90–103. [[CrossRef](#)]
17. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009.
18. Cosme, R.D.C.; Krohling, R.A. *Support Vector Machines Applied to Noisy Data Classification Using Differential Evolution with Local Search*; Technical Report; Universidade Federal do Espírito Santo: Vitória, Espírito Santo, Brazil, 2011.
19. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [[CrossRef](#)]
20. García Molina, J.F. Modeling and Analysis of Prostate Cancer Structures Alongside an Incremental and Supervised Learning Algorithm. Ph.D. Thesis, Ruprecht-Karls University of Heidelberg, Heidelberg, Germany, 2014.