P. Wefing, F. Conradi, J. Rämisch, P. Neubauer and J. Schneider

Determination of free amino nitrogen in beer mash with an inline NIR transflectance probe and data evaluation by machine learning algorithms

Free amino nitrogen (FAN) concentrations in beer mash can be determined with machine learning algorithms from near-infrared (NIR) spectra. NIR spectroscopy is an alternative to a classical chemical analysis and allows for the application of inline process quality control. This study investigates the capabilities of different machine learning techniques such as Ordinary Least Squares (OLS) regression, Decision Tree Regressor (DTR), Bayesian Ridge Regression (BRR), Ridge Regression (RR), K-nearest neighbours (KNN) regression as well as Support Vector Regression (SVR) to predict the FAN content in beer mash from NIR spectra. Various pre-processing strategies such as principal component analysis (PCA) and data standardization were used to process NIR data that were used to train the machine learning algorithms. Algorithm training was conducted with NIR data obtained from 16 beer mashes with varying FAN concentrations. The trained models were then validated with 4 beer mashes that were not used for model training. Machine learning algorithms based on linear regression showed the highest prediction accuracy on unpre-processed data. BRR reached a root mean square error of calibration (RMSEC) of 2.58 mg/L (R^2 = 0.96) and a prediction accuracy (RMSEP) of 2.81 mg/L (R^2 = 0.96). The FAN concentration range of the investigated samples was between approx. 180 and 220 mg/L. Machine learning based NIR spectra analysis is an alternative to classical chemical FAN level determination methods and can also be used as inline sensor system.

Descriptors: mashing, NIR, machine learning, FAN

1 Introduction

Free amino nitrogen (FAN) is an important quality parameter in the field of beer production [1–4]. It is generated during malting, extracted during mashing and usually analysed in the laboratory [3]. FAN compounds are the sum of di- and tripeptides, ammonium ions and amino acids [4]. As they are the primary nitrogen substrate for yeast during the later beer fermentation the FAN concentration is a major quality parameter in the mashing process. Adequate levels of FAN in wort ensure efficient yeast cell growth and, hence, a desirable fermentation performance [3–6].

In literature, varying FAN levels in beer mash are reported depending on the malt-liquor ratio and the used type of malts. Studies that used a comparable type of malt and malt-liquor ratio reported FAN values in the range of circa 170 mg/L – 320 mg/L [7–9]. Approximately up to 70 % of FAN is produced during malting, where higher nitrogen barley results in extracts that are carbohydrate rich and lower nitrogen barley results in extracts that are rich in carbohydrates [3, 10, 11]. During mashing, FAN is produced by proteolysis [8, 9, 12]. Mashing time and mashing temperature influence the levels of FAN produced during mashing [8, 9, 13].

Ninhydrin-based laboratory methods are commonly used for FAN analysis [3, 5, 14–17]. In FAN analysis, ninhydrin is added to a sample as an oxidizing agent that causes decarboxylation of α-amino acids. The reduced ninhydrin reacts with unreduced ninhydrin and liberates ammonia, forming a colour complex [17]. The FAN concentration is determined by measuring the intensity of the colour complex against a standard solution in a spectrophotometer. Additionally, high performance liquid chromatography (HPLC) methods are available for the analysis of amino acids in beer [18, 19]. HPLC methods rely on the ninhydrin method but allow for a separation of specific amino acids [3]. Alternatively, a gradient elution HPLC method using dansyl chloride as fluorogenic reagent is available for the quantification of individual amino acids in beer mash [3, 20].

In order to optimise the duration of the mashing process and the mash quality it is important either to determine the malt quality extensively prior to mashing or to measure mash quality during the process [8]. So far ninhydrin based methods do not allow for an inline FAN determination and are thus unlikely to be integrated

https://doi.org/10.23763/BrSc21-10wefing

Authors

Patrick Wefing, Florian Conradi, Johannes Rämisch, Jan Schneider, Technische Hochschule Ostwestfalen-Lippe, Institute of Food Technology, NRW, Lemgo, Germany; Peter Neubauer, Technische Universität Berlin, Department of Biotechnology, Bioprocess Engineering, Berlin, Germany; corresponding author: patrick.wefing@th-owl.de

into a process. Due to the significance of the FAN value as a quality parameter there is an increasing interest in the application of faster and easier applicable methods. A continuous solution for the prediction of FAN was already demonstrated by *Mitzscherling* et al. [8], who used an array of different bypass sensors such as pH, conductivity, ultrasonic velocity and viscosity. Multivariate statistical techniques (e.g. principal component regression, partial least squares regression) were used for the calibration of that bypass sensor array.

However, complex spectroscopic technologies, such as e.g. near-infrared (NIR) or Raman, have to our knowledge not been used for this analysis in an inline application. These methods are advantageous, as they only need a single sensor that can be implanted into the process. The application of NIR sensors has already been widely demonstrated in the fields of food and beer production [21, 22, 31–34, 23–30]. Apart from NIR, Fourier-transform infrared (FTIR) spectroscopy is also used for the laboratory analysis of FAN and soluble nitrogen content in the field of beer production [35, 36].

NIR spectroscopy is a non-invasive measuring method, allowing a quick real time analysis without the need for sample preparation [29]. Therefore, NIR is especially suited for the application in food and beer production. Used as an inline measurement system it allows the real time analysis of FAN during mashing and therefore enables a better understanding of the mashing process itself due to a process monitoring. Also, if available in real-time, the quality parameter FAN could be used as a control parameter during mashing for example by adapting the mashing time.

Photonic technologies such as NIR spectroscopy produce a large number of data and since the signals cannot be directly attributed to a specific substance, aside from common multivariate statistics approaches for the interpretation of NIR spectra as e.g. Partial Least Squares (PLS) regression, this is a field where machine learning algorithms may have a considerable potential [37]. PLS regression is established as the standard method for the analysis of NIR spectra [38]. However, PLS belongs to the field of classical statistics, aiming to formalise relations between independent and non-independent variables [39, 40]. In contrast to that, the aim of machine learning is the prediction itself than to find an explanation for the prediction. Here, inference is replaced by validation through testing the model with new data [39].

In this study, we investigate whether data from an inline NIR sensor, evaluated by machine learning algorithms, can be used for FAN level prediction in the field of beer production. Studies have revealed that machine learning algorithms such as Ordinary Least Squares (OLS), Ridge Regression (RR), Bayesian Ridge Regression (BRR), Decision Tree Regressor (DTR), K-nearest neighbours (KNN) regression, and Support Vector Regression (SVR) can be successfully applied to predict quality parameters from NIR spectra such as extract levels in beer, sugar levels in fruits, acidity of palm oil, moisture content of freeze dried products, and others [41–49].

NIR spectra contain a large number of indirect chemical information of the measured matrix. Often excessive data pre-processing steps are necessary to use the NIR spectra for quantification of single components. NIR pre-processing methods that are often applied individually or in combination are “first derivation of NIR spectra”, “multiplicative scatter correction”, “mean centring”, “spectra smoothing”, “Pareto scaling”, “standard normal variate” and “Savitzky-Golay filter” [43, 45, 50–53]. This complicates the industrial application of that technology and the inline implementation into processes. However, machine learning algorithms might be able to determine the FAN concentration in beer mash with a small or no data pre-processing step and high determination accuracy. Additionally, the available number of algorithms make the selection of an appropriate algorithm difficult.

This paper aims to compare different linear and non-linear machine learning regression strategies for the determination of FAN in beer mash. Mashes with different FAN contents were analysed by both, NIR spectroscopy and the ninhydrin assay as reference measurement. NIR spectra and reference data were then used to train machine-learning algorithms, which were validated with external data.

2 Materials and Methods

2.1 Mash Preparation

Finely ground malt (Pilsner malt, Weyermann, Germany) was used for the mashing experiments and demineralized water was used as brewing liquor (reverse osmosis). The malt was grinded with an impact mill (Millomat 100, Treffler Maschinenbau GmbH & Co. KG, Germany) at the Max Rubner-Institut -Federal Research Institute for Nutrition and Food in Detmold, Germany. The sieve used in that mill only allowed particles with a diamater $d_p \leq 500$ µm to pass. After mixing with the brewing liquor, the malt-liquor ratio was 1 : 4. The malt analysis of the used substrate is given in table 1.

Two stock mash batches (stock batch A and stock batch B) were prepared. Stock batch A consisted of a mash that was allowed to react for 30 min at 65 °C. The brewing liquor temperature for that batch before it was mixed with malt was 70 °C. It was heated up to 80 °C afterwards to inactivate enzymes and then cooled down rapidly to prevent further temperature induced degradation. Stock batch B was kept at 20 °C and was therefore not allowed to e.g. decompose starch into sugars.

Table 1 Malt analysis of Pilsner Malt

Type	Pilsner Malt
Producer	Weyermann (Germany)
Water Content	4.4 %
Extract Water Free	82.9 %
pH-value	5.88
Saccharification Time	10 – 15 min
Total Protein	10.6 %
Free Amino Acid	794 mg/100 g dry mass
Kolbach Index	41.4 %
Friability	88.4 %
Glassiness	1.2 %
Viscosity (referred to 86 g kg^{-1} Extract)	1.51 m Pa s

Subsequently the two stock batches were used to prepare 20 different mash batches with varying concentrations of FAN. The mashes were blended at 20 °C and each batch had a volume of approximately 1 L. The batch with the lowest FAN concentration consisted only of mash from stock batch b (0 % A : 100 % B) that was kept at 20 °C, while for the batch with the highest FAN concentration only mash from stock batch A was used (100 % A : 0 % B) . The other batches were blended using increasing concentrations of stock batch A (5 % – 95 %). Of the 20 prepared mash batches 16 were used for training of machine learning algorithms and four were used for validation purposes.

2.2 FAN Concentration Determination

FAN was measured using ninhydrin-based methods with the use of the absorbance measurement at 450 nm (CADAS 100, Hach Lange GmbH, Germany) according to the EBC-ninhydrin method by spectrophotometry [14, 17] .

2.3 Computation of Data

Python 3.7 was used for all calculations. The open source cross platform Spyder (MIT license) was applied as the programming environment. Machine learning libraries from scikit-learn 0.23 were utilized. A total of 1345 different NIR spectra were used in this study. Each spectra consists of 256 individual variables, describing the measured intensity at a different wavelength. The variables are referred to as features. Feature standardization was conducted by removing the mean and scaling the values to unit variance.

$$z = (X - \bar{X})/s \qquad \text{(Eq. 1)}$$

Here, z is the standard score of a sample, X is the value of a single feature and $\bar{X}$ is the average of the feature values. The dataset standard deviation is represented by s.

Principal component analysis (PCA) was also applied on NIR spectra. It is an unsupervised data dimension reduction technique frequently applied on large spectral data [43]. Unsupervised machine learning techniques are concerned with finding patterns and structures in unlabeled data, while supervised learning concerns learning from labeled data (e.g. NIR data that are trained with laboratory reference values as labels) [54, 55]. Each principal component (PC) is a linear combination of the original data. Data dimension reduction can lead to a simplification of the data set whilst maintaining a significant proportion of the variance, when the first few PCs can explain most of variance in the used data [56]. The first PC is the direction along which the samples show the largest variation, explaining the most amount of variance in the original data. The second PC is the direction that is uncorrelated (orthogonal) to the first component, but along which the samples show the largest variation [57].

2.4 Measurement of NIR Spectra

NIR spectra were obtained in a 1 L stainless steel stirred reactor (fill volume 500 mL) with a NIR process spectrometer (PSS 2120, Polyetc GmbH, Germany) equipped with a quartz halogen lamp. The mash was stirred during the measurement and the spectra were recorded in absorption mode at 20 °C. This differs from temperatures usually applied for the “protein rest” during mashing. In practice, a wide variety of different mashing procedures is found with different “protein rest” temperatures, that are e.g. 38 °C and 48 °C to 52 °C [58–60].

The enzymes involved in proteolysis are endo-proteinases, exopeptidases, and carboxypeptidases [12]. *Aldred* et al. demonstrated that no decrease in protein levels in beer mash occurs (due to enzymatic or physical protein degradation) below a temperature of 45 °C [12]. Therefore, no further change in FAN concentrations is expected at 20 °C. The chosen experimental setup (at 20 °C) thus allows for acquiring inline NIR sensor data and ninhydrin reference measurements at stable conditions.

The analysed wavelength range was 1100 – 2100 nm with a resolution of 3.9 nm. Baseline correction was conducted with demineralized water. An inline transflectance probe (Avantes BV, The Netherlands) was used with an optical path length of 10 mm (distance between optical fibre and reflecting mirror: 5 mm).

2.5 Statistical Evaluation

Model performance was judged by the root mean square error (RMSE) and R^2. The RMSE is used to measure the error of predictions by comparing the predicted value with the expected value and is frequently used for regression performance analysis [27, 41, 43, 44, 61–63]. This in contrast to the standard deviation, where the spread of data around a mean value is measured.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad \text{(Eq. 2)}$$

Here, y_i is the vector of the FAN levels determined with reference experiments (observed values) and $\hat{y}_i$ is the vector of FAN levels predicted with a machine algorithm (predicted values). Both, the training performance of the machine learning algorithm and the prediction performance of the trained model on independent and external validation data was evaluated with RMSE. RMSE for the training performance is referred to as root mean square error of calibration (RMSEC). The RMSE for the prediction performance of the trained model on validation data is referred to as root mean square error of prediction (RMSEP). Criteria to evaluate fitting accuracy are high R^2 and low RMSE values [43, 62].

The coefficient of variation (CV) was used to describe the repeatability and precision of the results obtained from FAN level prediction with NIR spectroscopy. The CV is calculated as the ratio of s to the corresponding mean value $\bar{x}$.

$$CV = \frac{s}{\bar{x}} \qquad \text{(Eq. 3)}$$

$$\text{with: } s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad \text{(Eq. 4)}$$

The observed values are $\{x_1, \ldots, x_n\}$ and $\bar{x}$ is the mean of the observed values. The sample size is demonstrated by the denominator N.

Also, the standard error (SE) indicating the distance from a sample mean to the mean of a population was calculated in this study for the FAN reference measurements.

$$SE = \frac{s}{\sqrt{n}} \quad \text{(Eq. 5)}$$

Here, s is the standard deviation and n is the numer of samples.

2.6 Machine Learning Strategy

This study aims to present an easy hands-on approach using machine learning for the determination of FAN in beer mash. FAN concentrations corresponding to raw NIR spectra of measured beer mashes were determined with laboratory experiments (reference data). The data were processed as shown in figure 1.

2.6.1 Raw Data

The used machine learning algorithms were trained on "training data" and validated on "validation data". Both datasets consist of NIR spectra, that were labelled with the corresponding FAN reference values obtained from the ninhydrin assay. From the 20 generated batches of mash, 16 were used for model training and 4 were used for model validation.

2.6.2 Data Pre-processing

The NIR data were prepared with four different pre-processing strategies that were used for independent approaches: (i) raw NIR data; (ii) standardized NIR data; (iii) dimensionally reduced NIR data by PCA (fist two components); (iv) standardized NIR data that were dimensionally reduced by PCA (first two components). Those steps were identical for training and validation data. Afterwards, the training data order was randomly changed (shuffled) and the data were split into a test and a training set (10 % test set, 90 % training set). A similar ratio for test and training set has already been reported in other studies [64–67]. Of note, the pre-processing does not influence the labelling of data, i.e. NIR spectra and corresponding FAN reference values are always shuffled as a pair.

2.6.3 Application of Different Machine Learning Algorithms

Machine learning algorithm training was then performed with the training set. Subsequently, the trained model was used to predict FAN values from the test set and the validation data.

2.6.4 Evaluation

To evaluate the prediction performance in comparison with the FAN reference values in terms of accuracy, the RMSEC and RMSEP was used.

2.7 Machine Learning Modules

Linear machine learning algorithms for regression are applicable if the target value is a linear combination of the individual features. In linear regression, the vector of observed values Y (dependent variable) is a linear function of the regressors X (predictor variable).

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{p1} \\ 1 & X_{12} & \cdots & X_{p2} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1n} & \cdots & X_{pn} \end{bmatrix}_{n\times p} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ \vdots \\ w_n \end{bmatrix}_{p\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1} \quad \text{(Eq. 6)}$$

Here, *w* is the vector of the unknown population to be estimated. Factors influencing Y other than X are referred to as error and are represented by ε. The FAN values are represented by Y and

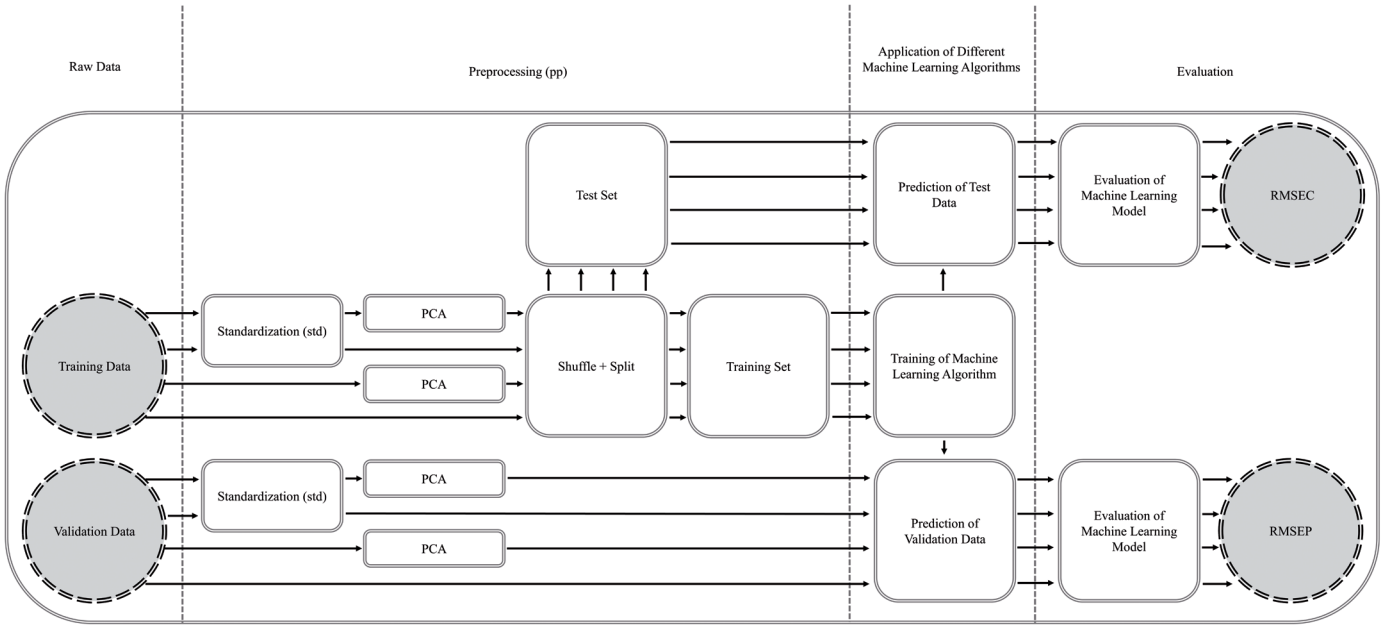

Fig. 1 Data processing steps for the determination of FAN content in beer mash. "Training Data" and "Validation Data" consist of NIR spectra and FAN reference values from the ninhydrin assay. Pre-processing such as standardization and dimension reduction with a PCA were only applied on the NIR data

the NIR data with 256 channels are represented by X (n × 256). Equation 6 can be rewritten as

$$Y = X\,w + \varepsilon \quad \text{(Eq. 7)}$$

In this work we aim to (i) train a model on a vector of reference FAN values Y and the corresponding NIR data matrix X by estimating the parameter vector $\hat{w}$. Then, we aim (ii) to predict FAN values from NIR spectra measured inline. The criteria used to determine $\hat{w}$ is to minimize the residual $\hat{\varepsilon}$.

$$\hat{\varepsilon} = Y - X\,\hat{w} \quad \text{(Eq. 8)}$$

The sum of squared residuals (RSS) is used to minimize $\hat{\varepsilon}$.

$$RSS = \hat{\varepsilon}^T\,\hat{\varepsilon} = (Y^T - \hat{w}^T X^T)\,(Y - X\,\hat{w}) \quad \text{(Eq. 9)}$$

Here, $\hat{\varepsilon}^T$ has the dimension 1 × n and $\hat{\varepsilon}$ has the dimension n × 1. The calculation of the estimator vector $\hat{w}$ by OLS is then given by

$$\hat{w}_{OLS} = (X^T X)^{-1} X^T Y \quad \text{(Eq. 10)}$$

$\hat{w}_{OLS}$ describes the estimator vector of the regression coefficients achieved by OLS regression and X is the matrix consisting of the features (NIR predictor variables). FAN reference values are the observed values Y.

RR is an alternative to OLS. It allows regularisation of the estimator $\hat{w}$ by introducing the penalizing scalar parameter a [47].

$$\hat{w}_{ridge} = (X^T X + a\,I)^{-1} X^T Y \quad \text{(Eq. 11)}$$

Here, I is the identity matrix. The application of a regularisation parameter works by trading increased bias for reduced variance [68], meaning that regularisation parameters reduce errors occurring during function fitting (training) and avoid overfitting. The bias represents the best possible fitting of an algorithm to a test data set of infinitive size [54]. Therefore, a low bias represents a relatively high correlation to the training data, while a high bias represents a relatively low correlation to the training data [54]. If the variance of a model to validation data is high while the bias to the test set is low, overfitting occurs.

Overfitting results in a minor predictive capability of a machine learning algorithm and often appears if a relatively small set of data is used for model training, as noise in that relatively small amount of data might be interpreted as real information [54]. This is of particular interest in the field of beer/food production, as the measurement of reference data is often time consuming and the number of data are therefore limited.

The regularisation parameter aims to prevent overfitting by increasing the bias to the test data set and decreasing the variance from generated model to the validation data set at the same time, changing the bias/variance trade-off of a machine learning model. If a = 0, the regularisation parameter shows no effect on RR and the estimation results will be equal to OLS. Choosing a regularisation parameter a > 0 mimizes $\hat{w}$ and aims to reduce the error on generalized or validation data, while the bias on OLS [47].

An estimator that shows a relatively small bias on the training data and has a higher RMSE and/or R^2 score than an unbiased estimator is closer to the real value of the parameter [46]. If finally the variance of the ridge estimator $\hat{w}_{ridge}$ could be reduced, the mean squared error tends to be lower compared to OLS [46].

BRR is a variant of RR that formulates linear regression problems by the aid of probabilities.

$$y \sim N(w\,X, \sigma^2) \quad \text{(Eq. 12)}$$

Here, the output vector y is viewed as variable of which the elements (w^T, X) are distributed in a Gaussian distribution $N(\ldots)$. The variance of the distribution is given by σ^2. According to Bayes' theorem, the posterior probability of the parameter w is proportional to likelihood and prior.

$$Posterior = \frac{Likelihood \cdot Prior}{Evidence} \quad \text{(Eq. 13)}$$

The evidence term is commonly ignored in model fitting with Bayes regression [69]. The posterior is defined as the revised probability of an occurring event depending on the available data.

$$Posterior = P(w|y, X, a, \sigma^2) \quad \text{(Eq. 14)}$$

Here, a is the parameter used to impose a penalty on the size of the coefficient w similar to RR. Likelihood represents the distribution of the observed data in relation to its parameters as probability.

$$Likelihood = P(y|X, w, a) = N(y|w\,X, \sigma^2) \quad \text{(Eq. 15)}$$

The output y is assumed to form a Gaussian distribution around $w \cdot X$. The prior is the parameter distribution before the data are observed.

$$Prior = P(w|\sigma^2) = N(0, \sigma^2) \quad \text{(Eq. 16)}$$

Estimation of the coefficient w occurs during model fitting, while the (regularisation) parameters a and σ^2 being updated by the "log marginal likelihood" function as described by *Tipping* et al. [70].

Apart from the linear machine learning strategies mentioned above, non-linear algorithms also exist. As demonstrated before, also DTR is used for analysis of NIR spectroscopy data in this work [42]. Building a decision tree (in terms of creating nodes) is based on decisions (criteria) that are arranged in a flowchart-like structure. Inside that structure nodes represent an attribute test (e.g. whether a value is above or under a certain threshold). As a result of that test, new nodes are created. The different node paths are referred to as branches. The terminal end of the decision tree structure are leafs. The whole path beginning from the root node and ending in the leaf represents the prediction rules for the trained model.

However, various decision tree algorithms are available. In this work, the classification and regression tree (CART) algorithm that creates nodes based on binary decisions was used for model training and FAN value prediction from NIR data. Growing trees in CART implies recursive splitting of tree nodes into a left and a

right branch: a root node containing all the calibration data and two children nodes (left and right) containing data observations [42].

The DTR uses the mean squared error (MSE) to decide how to split a node in sub-nodes. CART algorithm divides training data depending on the attributes k (category) and t_k (threshold for category). Determination of k and t_k is accomplished by the loss function $J(k, t_k)$:

$$J(k, t_k) = \frac{s_{left}}{s} MSE_{left} + \frac{s_{right}}{s} MSE_{right} \quad \text{(Eq. 17)}$$

with: $$MSE_{node} = \sum_{i \in node} (\hat{y}_{node} - y_i)^2 \quad \text{(Eq. 18)}$$

and: $$\hat{y}_{node} = \frac{1}{s_{node}} \sum_{i \in node} y_i \quad \text{(Eq. 19)}$$

Here, s is the subset of features and the mean squared error (MSE) is used as optimization indicator.

The KNN algorithm was also used for the prediction of FAN levels from NIR data. KNN allows for a non-linear observation of coherences describing the relation between the NIR data and the FAN reference values [71]. To predict values (e.g. FAN levels) from any new data (e.g. NIR spectra), similarities between the training data (e.g. FAN concertation and NIR spectra) and the data used for prediction are determined. Therefore, prediction of a new value is based on how closely the measured data (NIR spectra) resemble the set of training data. The constant k is used to define the numbers of neighbours used for the prediction. Here, the nearest five neighbours (k = 5) were used, meaning that a value was predicted from the five data points, with the most similar structure. To find the closest neighbours a distance function calculating the difference between two stances is used [43].

$$d_{manhatten} = \sum_{i=1}^{k} |o_n - o_t| \quad \text{(Eq. 20)}$$

In Equation 20, the distance between the vectors of a new observation o and an observation used for model training o_t is determined.

Non-linear regression based on Support Vector Machines (SVM) was also tested. It is referred to as SVR [72]. SVM is based on the calculation of hyper-planes in the input feature space to separate classes with maximal margin. The closest samples to the decision boundary defining the hyper-plane position are called support vectors [1].

SVR uses a penalty parameter (ε_{SVR}) which allows for a selection of prediction parameters within a certain range only. That channelizes the available data into a tube, whereas small ε_{SVR} values allow for a reduced error tolerance and vice versa [55]. SVR formulates a function approximation problem as an optimization problem that attempts to find the narrowest tube possible centred around the available data [55]. The support vectors used for prediction are located at the boundary of the area defined by ε_{SVR}.

However, SVR can apply a "kernel trick", that allows the algorithm to operate in a higher dimension without computing the coordinates in that space. This can help to solve non-linear problems [73]. In principal, the input data X can be transformed (ϕ maps) into a higher m-dimensional feature space: $\phi: X \rightarrow \phi(X)$. E.g. a positive definite kernel K is expressed as

$$K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle \text{ with } x_i, x_j \in X \quad \text{(Eq. 21)}$$

Then non-linear problems can be solved linearly [1, 73].

$$f(X, w) = \sum_{i=1}^{m} w_i \, \phi_i \, (X) + b \quad \text{(Eq. 22)}$$

Here, b is the bias term. w_i and ϕ_i are the coefficients and non-linear transformations, respectively [1]. Commonly used kernel functions that were also applied in this study are the linear kernel and the Radial Basis Function kernel (rbf) [73].

linear kernel: $K(x_i, x_j) = \langle x_i \cdot x_j \rangle$ (Eq. 23)

rbf: $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$ (Eq. 24)

with: $\gamma = \frac{1}{m \, \sigma_X^2}$ (Eq. 25)

The spread of the rbf was controlled by γ, where m is the number of features and σ_X^2 is the variance of X. The quality of the estimation is then measured by a loss function that is sensitive to ε_{SVR} [1].

$$min \; \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^{n} (\xi_i + \xi_i^*) \quad \text{(Eq. 26)}$$

such that: $w^T x_i - b + y_i \leq \varepsilon + \xi_i$ (Eq. 27)

and: $w^T x_i + b - y_i \leq \varepsilon + \xi_i^*$ (Eq. 28)

with: $\xi_i, \xi_i^* \geq 0$ (Eq. 29)

Here, n is the number of inputs, x_i and y_i are the inputs and targets, and ξ_i and ξ_i^* are the slack variables defining if a prediction lies above the space that is created by ε_{SVR}. The cost parameter $C > 0$ determines the trade-off between the size of w (a small w is desired) and up to which deviations larger than ε_{SVR} are tolerated [74]. In this work, C was set to 1. Further information regarding the application of SVR can be seen in the publication of *Smola* et al. [74].

3 Results

FAN levels in beer mash were analysed with NIR spectroscopy and machine learning algorithms. FAN reference values were obtained from a ninhydrin assay. Samples from 20 different mashes with varying FAN concentrations were prepared and analysed. 16 of those samples were used for model training, 4 for model validation. The results of the FAN ninhydrin assay analysis are shown in table 2 (see page 113).

Table 2 Results of the FAN nynhidryn assay reference analysis. n = 6

Batch No.	FAN [mg/L]	s [mg/L]	CV [%]	SE [mg/L]	Training	Validation
1	220.22	± 1.29	± 1.67	± 0.53	Yes	No
2	218.29	± 1.77	± 3.12	± 0.72	No	Yes
3	215.68	± 1.41	± 1.99	± 0.58	Yes	No
4	213.77	± 1.14	± 1.29	± 0.46	Yes	No
5	206.89	± 2.48	± 6.16	± 1.01	Yes	No
6	206.50	± 1.19	± 1.41	± 0.48	No	Yes
7	204.69	± 1.66	± 2.75	± 0.68	Yes	No
8	202.72	± 1.53	± 2.34	± 0.62	Yes	No
9	201.76	± 1.41	± 2.00	± 0.58	Yes	No
10	199.41	± 1.85	± 3.43	± 0.76	Yes	No
11	199.13	± 2.30	± 5.30	± 0.94	No	Yes
12	197.12	± 1.72	± 2.95	± 0.70	Yes	No
13	194.81	± 1.12	± 1.26	± 0.46	Yes	No
14	192.48	± 1.72	± 2.94	± 0.70	Yes	No
15	190.63	± 1.91	± 3.64	± 0.78	Yes	No
16	190.62	± 1.33	± 1.77	± 0.54	Yes	No
17	187.15	± 0.88	± 0.78	± 0.36	Yes	No
18	186.24	± 1.44	± 2.08	± 0.59	No	Yes
19	183.00	± 2.26	± 5.09	± 0.92	Yes	No
20	181.86	± 0.97	± 0.94	± 0.40	Yes	No

The laboratory analysis showed that the FAN concentration of the investigated mash samples was in a range of 181.86 mg/L (s = ± 0.97 mg/L) and 220.22 mg/L (s = ± 1.29 mg/L).

3.1 Spectra Presentation

The NIR spectra used for training of machine learning algorithms are displayed in figure 2. Areas with high noise levels were detected between 1400 nm – 1500 nm and 1900 nm – 2100 nm in the raw data sets. Those areas also clearly show noise in the standardized spectra.

The noise areas correspond to the first and second water overtone areas in NIR spectra. Areas before and between the noise areas are distinguishable in standardized and raw NIR data. FAN level and absorbance do not correlate. The position of the spectra in the figure 2 shows a randomized behaviour rather than an organized distribution induced by different FAN levels. A determination of FAN levels is not possible from NIR spectra in this processing status.

3.2 Principal Component Analysis of NIR Spectra

A dimensionality reduction of data was conducted with a PCA. The data were transferred from a space with 256 dimension to a space with 2 dimensions (Fig. 3, see page 114). PCA of raw NIR data showed an explained variance of 13.94 % for the first and 5.93 % for the second principal component. Standardized data decomposition resulted in variances explaining 55.51 % (PC 1) and 20.77 % (PC2).

The PCA of raw NIR data does not permit the separation into independent clusters. However, it is visible that data representing a lower FAN are oriented in the lower left segment of figure 3A and those for medium and higher FAN in opposite direction. A separation between the different concentrations is still not possible as the sections (low, medium, high FAN) overlap each other.

The PCA of standardized data resulted in fairly separable clusters. The clusters only allow a clear separation of spectra with relatively high differences in FAN concentrations as shown in figure 3 B. However, data representing the highest FAN levels (darkest sports) are clustered on the right and the left side of the plot (Fig. 3 B), indicating that FAN levels are not the main driver of separation of the observed clusters in the direction of PC 1.

Other factors influencing the quality of the NIR spectra such as gradually dropping absorbance levels during measurement were not observed in this study.

3.3 Training of Machine Learning Algorithms

Machine learning algorithms were applied on the NIR spectra

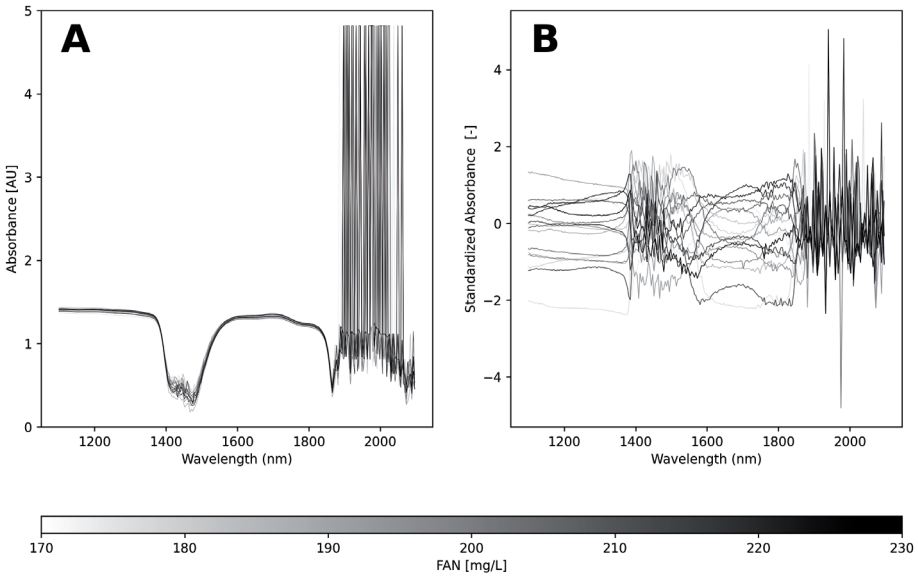

Fig. 2 Example of NIR spectra used for training of machine learning algorithms. A: raw NIR spectra, B: standardized NIR spectra; For a better clearness only around 15 % randomly chosen of the data are shown in this figure. FAN levels of the spectra are indicated by the grey scale

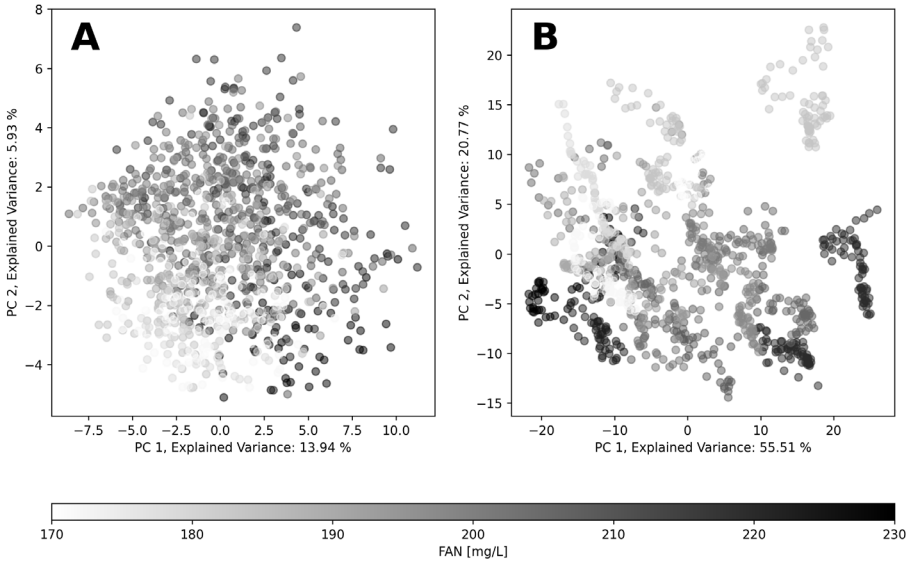

Fig. 3 PCA of (A) raw NIR data and (B) standardized NIR data. 100 % of the data are shown. FAN levels of the individual spectra are displayed by the grey scale

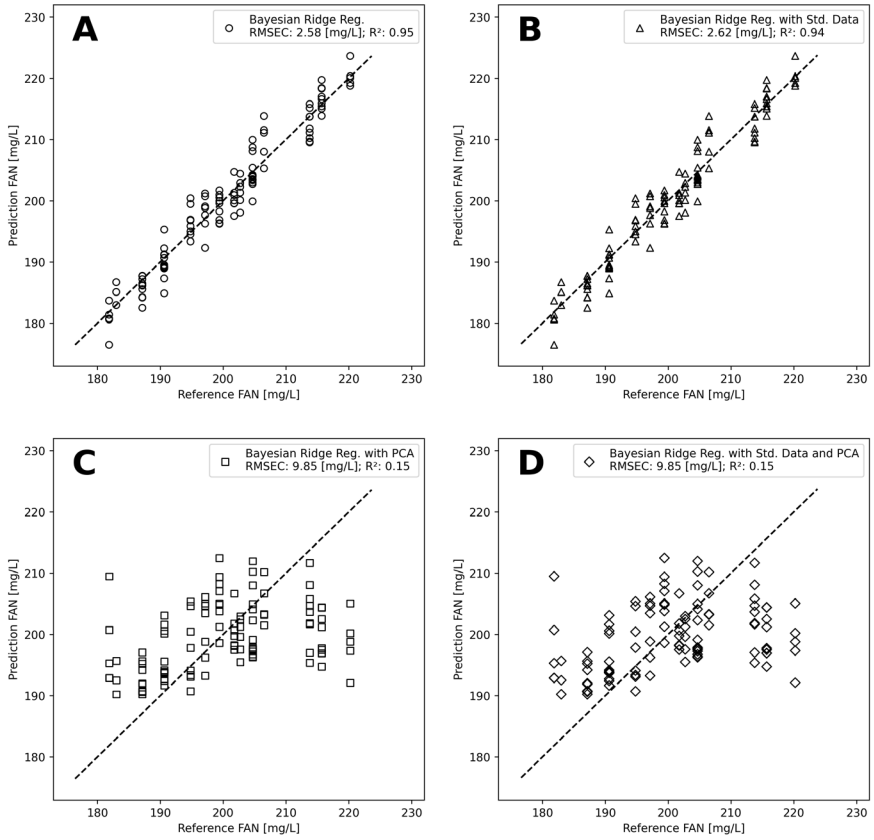

Fig. 4 Training results with test data of BRR for the prediction of FAN from NIR data. A: without pre-processing of training data ; B: with standardization of training data before model training; C: with PCA of training data prior to model training; D: with standardized data used for PCA prior to model training. The dashed line indicates an ideal fit between predicted and reference values

gathered from mash samples. The best training results for raw NIR spectra were obtained for OLS, DTR, RR and BRR. The complete results of the investigated machine learning algorithms are shown in table 3, see page 115.

Training results of BRR and RR applied on raw data showed the best fits relating to the RMSEC and R^2 scores. The results are shown in figure 4. As indicated by the obtained scores (Table 3), training with raw and standardized data show a similar distribution of predicted values (Fig. 4 A and B). Application of data pre-processed with PCA and a combination of PCA and standardization showed a lower accuracy compared to the raw data and standardized data for the BRR approach and for all the other algorithms tested. Both the results of fitting with PCA and a combination of PCA and standardized data showed a similar distribution as displayed in figure 4 C and D as expected given the similar score values in table 3.

The RR algorithm can be regularised by the scalable parameter value α (cf. Eq. 11) to achieve a potentially lower bias on the training set and a greater variance on the test set. A high bias is prone to overfitting of the trained model. In comparison with OLS, where the bias cannot be controlled by an additional variable imposed on the learning coefficients, only slightly better RMSEC values were achieved with an value of $\alpha = 1 \cdot 10^{-6}$. Also, α values of $1 \cdot 10^{-2}$ and 1 were tested on the data, but with worse prediction accuracy compared to an α value of $1 \cdot 10^{-6}$ (cf. Table 3).

Improved training results with standardized data compared to raw data were found for SVR (rbf and linear kernel), DTR and KNN (k = 5).

Training of data with PCA pre-processing showed no RMSEC values below 9.70 mg/L, regardless if the NIR data were standardized before decomposing, or not.

Overall model training with DTR and standardized data provided the best results in terms of the lowest RMSEC. The training results are shown in figure 5, see page 115. Again, data pre-processed with PCA and a combination of PCA and standardization did not allow a sufficient FAN prediction in the training stage and were therefore not considered for validation on external data.

3.4 External Validation of Trained Machine Learning Models for the Determination of FAN

The trained machine learning models were validated with data not included for the model training. Those data were provided by independent experiments representing external data to the training

Table 3 Results of machine learning training; Range of investigated FAN values: approx. 180 mg/L to 220 mg/L

Algorithm	without pre-processing		with standardization		with PCA		with PCA and standardization	
	RMSEC [mg/L]	R^2	RMSEC [mg/L]	R^2	RMSEC [mg/L]	R^2	RMSEC [mg/L]	R^2
OLS	2.60	0.95	2.60	0.95	9.85	0.15	9.85	0.15
SVR								
RBF	7.72	0.50	4.63	0.83	9.70	0.19	9.70	0.19
Linear	5.91	0.69	3.11	0.92	9.99	0.15	9.99	0.15
DTR	3.44	0.90	2.13	0.96	13.28	0.06	13.78	0.03
RR								
$\alpha = 1 \cdot 10^{-6}$	2.58	0.95	2.60	0.95	9.85	0.15	9.85	0.15
$\alpha = 0.01$	3.08	0.92	2.59	0.95	9.85	0.15	9.85	0.15
$\alpha = 1$	4.76	0.80	2.75	0.94	9.85	0.15	9.85	0.15
KNN								
k = 5	8.30	0.40	2.51	0.98	10.04	0.17	10.04	0.17
BRR	2.58	0.96	2.62	0.94	9.85	0.15	9.85	0.15

stage. The validation results are shown in table 4, see page 116.

As seen already for the (internal) training evaluation OLS, BRR and RR showed better results in terms of lower RMSEP values (compared to the other three model types). However, the DTR showed comparably high RMSEP values in the external validation whereas in the training status a relatively low RMSEC value was found. With standardized NIR data DTR allowed the FAN prediction with the lowest RMSEP of all tested algorithms. For all other algorithms, models with standardized data showed higher RMSEP values compared with unpre-processed data. The correlation of predicted FAN concentrations to the laboratory reference data was best with BRR when using raw (unpre-processed) NIR data, as shown in figure 6, see page 116.

The lowest CV was found for the BRR model that was 1.03 % (220.73 mg/L ± 2.26 mg/L), while the highest 1.26 % (200.33 mg/L ± 2.52 mg/L) indicating a good reproducibility of the regression model according to the requirements of industrial practice (cf. Fig. 6). Predicted FAN levels from 64 NIR measurements were used to calculate the coefficients of variation, each. The lowest RMSEP value for validation with standardized data was 7.53 mg/L found for DTR.

4 Discussion

In this study, the application of NIR spectroscopy in combination with an inline probe for FAN quantification in beer mash was shown for the first time. NIR spectroscopy allows for a chemical feedback based on infrared light absorption. Mitzscherling et al. already have demonstrated the application of a sensor array for FAN online monitoring using a bypass [8]. However, the then used system concentrated on rather physical measurement parameters such as ultrasonic and conductivity measurements

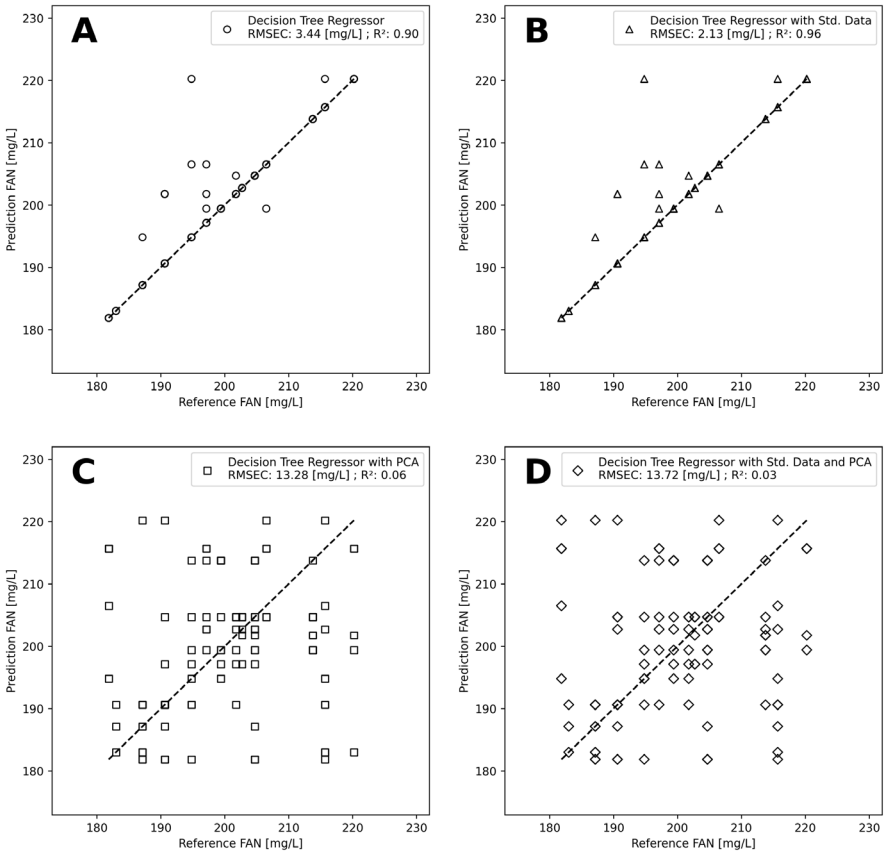

Fig. 5 Training results with test data of DTR for the prediction of FAN levels from NIR data. A: without pre-processing of training data; B: with standardization of training data before model training; C: with PCA of training data prior to model training; D: with standardized data used for PCA prior to model training. The dashed line indicates an ideal fit between predicted and reference values

Table 4 **Model validation on external data; RMSEP values (range of FAN values: approx. 180 mg/L to 220 mg/L) representing the accuracy of the machine learning model as deviation from the original value. For RMSEP values, the lower the value the better the prediction**

Algorithm	without pre-processing		with standardization	
	RMSEP [mg/L]	R^2	RMSEP [mg/L]	R^2
OLS	3.04	0.95	21.32	0.04
SVR				
RBF	9.44	0.40	10.24	0.25
linear	7.81	0.58	12.15	0.21
DTR	11.17	0.12	7.53	0.79
RR				
$\alpha = 1 \cdot 10^{-6}$	3.00	0.95	21.32	0.04
$\alpha = 0.01$	3.19	0.95	21.35	0.04
$\alpha = 1$	4.94	0.84	17.63	0.01
KNN				
k = 5	10.08	0.37	12.86	0.05
BRR	2.81	0.96	20.03	0.02

reaching an accuracy (RMSEP) of 15.60 mg/L FAN using PLS regression. In our current work, we evaluated the combination of NIR measurements with machine learning algorithms. The best FAN prediction models ($R^2 \geq 0.95$) were BRR, RR ($\alpha = 1 \cdot 10^{-6}$), OLS, and RR ($\alpha = 0.01$) with 2.81 mg/L, 3.00 mg/L, 3.04 mg/L, and 3.16 mg/L, respectively.

Given the above, the FAN prediction performance of the NIR machine learning setup seems to be more advantageous compared to the sensor array setup presented by Mitzscherling et al. [8]. In addition, for inline NIR measurement only a single senor is needed, simplifying the technical implementation.

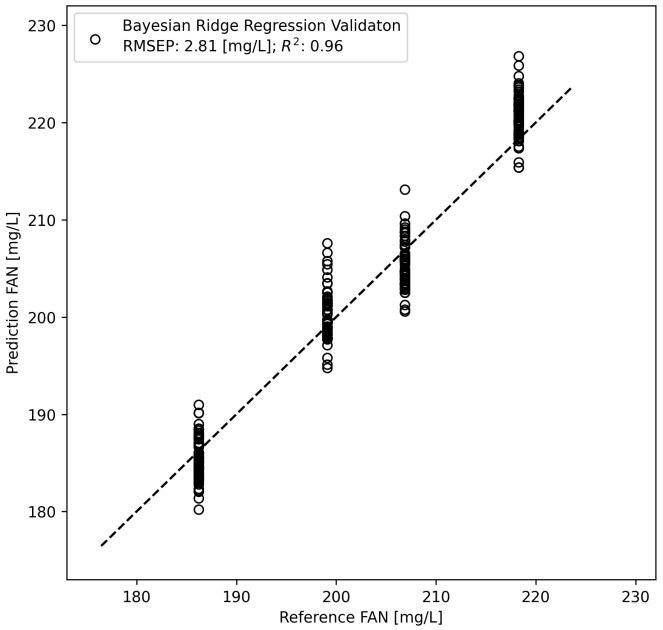

Fig. 6 **Validation results of BRR model with external data for the inline prediction of FAN without pre-processing of training data. Dashed line indicate an ideal fit between predicted and reference values. Each cluster consists of 64 values.**

4.1 Precision and robustness of the NIR FAN determination in comparison to the reference method

Reference laboratory data are the foundation for supervised machine learning methods. Supervised learning relies on prior knowledge about an example dataset to make predictions about new data points, e.g. predicting the value of a response variable on the basis of the input variables [75]. The quality and accuracy of those data is essential for the machine learning performance. The ninhydrin assay for the determination of FAN levels in beer used in this paper was published by Lie and is the EBC/MEBAK standard method [17]. Analysing 14 replicates Lie obtained a variation coefficient of 4 – 6 % in FAN level analysis [17].

The majority of FAN ninhydrin assay results obtained from laboratory experiments in this study were below a variation coefficient of 6 %, with an exception for the variation coefficient of a single batch that was 6.61 %. In this study, 6 replicates were used. Compared to Lie higher coefficients of variation were expected, since the CV decreases with increasing sample size [76]. However, the obtained variation coefficients were similar or lower than the previously reported numbers proving that the employed assay has a high level of precision and repeatability.

Correspondingly, the precision of the results from the ninhydrin assay used for the reference analytics in this work measured by the standard deviation was higher compared to the results published by Lie [17]. The quality of the reference values is also demonstrated by the relatively low SE values, showing that the calculated mean values are close to the theoretical mean values of the population. A satisfactory RMSEP and R^2 indicate that sample amount and quality (precision and repeatability) were sufficient for an accurate machine learning based prediction model. Unsatisfying RMSEP values might be due to reference methods unable to perform measurements in certain level of repeatability and precision or due to an incongruous choice of machine learning algorithms.

Prediction of FAN levels from NIR spectra with BRR resulted in even lower coefficients of variation than observed for the majority of reference measurements (18 of 20). The reason for this is the greater number of samples used for the validation measurement. 64 individual NIR spectra were taken from each of the four batches used for validation, while 6 measurements were conducted to determine FAN levels with the ninhydrin assay. As mentioned above, the CV is unproportional to the number of samples and decreases with increasing sample number [76].

Accuracy and precision measured by RMSEP (and R^2) and the CV suggest that the NIR inline setup is able to predict FAN levels in a satisfying quality. However, the precision also depends on the number of samples used to predict a certain value (cf. determination of the CV). In this study, 64 for samples in the validation stage were used to calculate the CV. A lower number of samples might lead to a lower CV value.

The question whether an analytical method is robust or which

factors possibly compromise the robustness of a method will be important if it comes to its application. The definition of robustness is similar in the field of chemical analysis and machine learning. In the field of chemical analysis, robustness of an analytical procedure is a measure of its capacity to remain unaffected by variations in the method parameters [77, 78]. Examples of such variations are the stability of the analytical sample and temperature fluctuations.

In the field of machine learning, robustness means that an algorithm delivers trustworthy output values even if the input data for model training or data used for prediction from a model are e.g. affected by some perturbations [79–81]. Among many reasons for altered data are the precision of measurement instruments, quantization errors and the presence of noise [79]. Furthermore, a robust machine learning model is characterized by the reliable prediction of values from new and unseen data that are similar to the data used for training. Equally, both definitions of robustness emphasise that a robust procedure is unaffected by disturbances.

Beer mash is composed of malt, a substance of natural origin. Biological variations of the mash are unavoidable and will therefore play a role in the industrial application of inline FAN NIR determination. Noise areas in the NIR spectra between 1400 nm – 1500 nm and 1900 nm – 2100 nm resulted in a perturbation of the NIR data. Furthermore, the quality of the FAN reference measurements are crucial for the machine learning models, as those results are used as input data for the training algorithms. Yet, the comparison of BRR training (RMSEC = 2.58 mg/L) and validation (RMSEP = 2.81 mg/L) metrics demonstrated, that the investigated method is able to perform inline FAN prediction of a trustworthy robustness. Training and validation results of RR and OLS indicated a similar level of robustness.

The machine learning algorithms were trained within a range of approximately 180 mg/L – 220 mg/L. Therefore, the prediction of FAN values are expected to have a high accuracy, precision and stable robustness within the investigated FAN range.

The inline measurements in this study were conducted at 20 °C. However, typical “protein rest” temperatures are e.g. 38 °C and 48 °C to 52 °C [58–60]. The chosen temperature prevented a change of the investigated mash during the measurement. This allowed for a judgement of the FAN determination performance on a stable measurement matrix. At temperatures that are usually employed for the “protein rest”, FAN levels would increase during the measurement. This might have an influence on the accuracy and robustness of the inline FAN determination by NIR.

4.2 Quality and speed of NIR FAN determination allows for an industrial inline application

Quality parameters such as FAN can potentially be used as a control parameter in an industrial process. High FAN prediction performance and fast inline data availability can be provided by NIR spectroscopy. In this study, the measurement time for one NIR spectrum was 640 ms. The choice of algorithm selection for data processing plays a deciding role for the exploitation of inline NIR for FAN determination.

In terms of achieving a fast and reliable inline FAN determination with NIR, linear regression models without the application of data pre-processing showed the greatest potential in this study. Elaborate pre-processing is time consuming and prone to errors.

Apart from fast measurement times, high accuracy is required for industrial sensor systems. Here also, the linear regression models demonstrated the most accurate prediction performance and therefore a high potential for the industrial application. They allow inline FAN determination from NIR data in the range of accuracy of the reference laboratory method. Among the linear regression algorithms, the Bayesian method showed the best prediction performance, which is at least partly explainable.

RR and OLS showed only slightly worse performances. However, OLS prevents utilization of regularization parameters and is therefore strongly dependent on data quality and data pre-processing. RR allows regularisation parameters but only to a limited extend, to change the bias of the training data set. BRR in contrast provides variable regularization parameters that can even further adapt to a given data set allowing it to exceed the limits of the RR regularisation parameters.

4.3 BRR allows for small datasets due to its flexibility

Concerning the practical application in industrial food and beverage production, the limited amount of data is generally an issue as laboratory reference analysis are usually costly. Adaptability on the circumstances in a production site is therefore necessary for an application in an industrial process and that can be best achieved by the application of the comparable flexible BRR.

BRR is applicable for use on small sets of data [82]. It is flexible with its choice of prior density assigned to measurement related data variabilities, and it allows the implementation of models that estimate shrinkage and perform variable selection [83]. BRR application on data sets of varying sizes collected with measuring instruments of varying precision can still lead to satisfactory prediction models due to the flexibility of the algorithm itself.

4.4 DTR showed overfitting

DTR performance on training data tend to show a misleading indication of the algorithm’s true predictive capacity [65]. In this study, DTR showed a high accuracy on training data and a far worse performance on validation data, indicating an overfitting. That probably resulted from the fitting of noise and peculiarities in training data rather than finding a general predictive rule [84]. Much effort on the general design of the decision tree would be necessary for better and perhaps acceptable prediction result. However, the costly and time-consuming preparation of such an elaborated and sensible design is not suitable for an industrial inline application requiring a certain degree of robustness.

4.5 Boundary conditions such as interaction of water and NIR are unavoidable in mash

Boundary conditions are circumstances that are uninfluencable e.g. due to technical limitations of the analytical equipment or the

composition of a sample. In this study, the main component of the mash samples analysed with NIR spectroscopy was water. Usually, the malt-liquor ratio during mashing is 1 : 4 or 1 : 3 [58, 85–88]. In this study, a ratio of 1 : 4 was used, whereby the influence of water on the NIR spectra was expected to be greater than of a mash with a malt-liquor ration of 1 : 3 due to the higher amount of water.

The NIR spectra obtained from mash samples showed two noise areas in the first overtone region and in the combination band region of water. The noisy areas were induced by the relatively high water content of the mash samples. They resulted in artefact values leading to reach the maximum capacity of the detector, especially in the combination band region and have therefore a deteriorating effect on the NIR data quality.

However, the high water content of the mash samples cannot be avoided in an inline measurement. Therefore, a suitable machine learning method for inline FAN determination from NIR data must be able to adapt to the noisy areas either by the algorithm itself or by the application of pre-processing methods.

4.6 Standardization of data

Data pre-processing, in general, has the potential to significantly enhance the results of machine learning algorithms in terms of accuracy [89]. On the other hand, pre-processing methods are a complication of machine learning strategies due to an increasing number of processing steps and can impair the results. A low number of steps needed for a machine learning strategy might lead to a less error prone and therefore more robust concept [55]. Therefore, only cost efficient pre-processing strategies such as standardization and dimension reduction by PCA were applied in this study.

Data standardization is one such a cost efficient method, that is widely used in the field of machine learning and often used to improve training performance [90]. The anticipated effect of utilizing data standardization is to change the ratio between relatively high values from noisy areas of the NIR spectra and relatively low values from non noisy areas towards more balanced values compared to the raw data.

The standardization of data led to improved or similar training accuracies for the tested algorithms compared to training on raw data. Nevertheless, overfitting was observed for all algorithms trained with standardized data at the validation stage. Therefore, standardization of NIR data was demonstrated to be unsuitable as data pre-processing method in this study.

4.7 Dimension reduction via PCA

The explained variances obtained from PCA of raw data indicate that only a small amount of the original information of the data is included in the dimensionally reduced dataset. Furthermore, a separation of the data into clusters did not occur. A dimension reduction with PCA aims to extract the important information concerning the variation in the data, to represent them as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables [91, 92].

As no patterns of similarity were found, PCA is probably not suitable as data pre-processing step for the data obtained in this study. The areas of noise in the NIR spectra probably have contributed to that. Removing those areas and/or the application of spectral pre-processiong steps such as Savitzky-Golay filter or smoothing of the spectral data might be necessary if PCA is used. Accordingly, PCA as a pre-processing strategy was not implemented for the validation with external data.

5 Conclusion

NIR spectroscopy with an inline transflectance probe is able to predict FAN levels in beer mash by the application of machine learning algorithms. A variety of six different machine learning regression algorithms were tested on mashing samples. The tested machine learning models utilized linear and non-linear algorithms.

Linear algorithms showed the highest accuracy for the inline prediction of FAN levels in beer mash in this study. Among them, BRR showed the most promising results in terms of accuracy and precision. Different data pre-processing steps from the field of machine learning were tested to increase the performance of the tested algorithms. Neither of them contributed to an increased precision and/or accuracy of the tested models. Omitting pre-processing steps contributes to a reduced effort in the implementation of that technology.

Acknowledgments

Financial support by the Federal Ministry of Education (BMBF), project 13FH3I01IA, in the frame of smartFoodTechnologyOWL, and the Ministry of Culture and Science of North Rhine-Westphalia is gratefully acknowledged.

Declaration of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6 References

1. Cernuda, C.; Lughofer, E.; Klein, H.; Forster, C.; Pawliczek, M. and Brandstetter, M.: Improved quantification of important beer quality parameters based on nonlinear calibration methods applied to FT-MIR spectra, Analytical and Bioanalytical Chemistry, **409** (2017), no. 3, pp. 841-857.
2. Choi, E.-J.; Park, J.-W.; Kim, J.H. and Kim, W.J.: Biological acidification and beer quality: addition of lactic acid bacteria isolated from malt, Journal of the Institute of Brewing, **126** (2020), no. 2, pp. 176-183.
3. Hill, A. and Stewart, G.: Free Amino Nitrogen in Brewing, Fermentation, 5 (2019), no. 1, p. 22.
4. Stewart, G.: Saccharomyces species in the Production of Beer, Beverages, **2** (2016), no. 4, p. 34.
5. Lekkas, C.; Stewart, G.G.; Hill, A.; Taidi, B. and Hodgson, J.: The Importance of Free Amino Nitrogen in Wort and Beer, Technical Quarterly-Master Brewers Association of the Americas, **42** (2005), no.

2, pp. 113-116.
6. Estela-Escalante, W.D.; Rosales-Mendoza, S.; Moscosa-Santillán, M. and González-Ramírez, J.E.: Evaluation of the fermentative potential of Candida zemplinina yeasts for craft beer fermentation, Journal of the Institute of Brewing, **122** (2016), no. 3, pp. 530-535.
7. Jacob, F.F.; Hutzler, M. and Methner, F.-J.: Comparison of various industrially applicable disruption methods to produce yeast extract using spent yeast from top-fermenting beer production: influence on amino acid and protein content, European Food Research and Technology, **245** (2019), no. 1, pp. 95-109.
8. Mitzscherling, M.; Becker, T.; Delgado, A.; Back, W.; Krottenthaler, M. and Kühbeck, F.: Online Monitoring of Gravity, FAN and β-Glucan during Mashing, Journal of the Institute of Brewing, **113** (2007), no. 3, pp. 293-301.
9. Kühbeck, F.; Dickel, T.; Krottenthaler, M.; Back, W.; Mitzscherling, M.; Delgado, A. and Becker, T.: Effects of mashing parameters on mash β-glucan, FAN and soluble extract levels, Journal of the Institute of Brewing, **111** (2005), no. 3, pp. 316-327.
10. Jones, B.L. and Budde, A.D.: How various malt endoproteinase classes affect wort soluble protein levels, Journal of Cereal Science, **41** (2005), no. 1, pp. 95-106.
11. Agu, R.C.: Some Relationships Between Malted Barleys of Different Nitrogen Levels and the Wort Properties, Journal of the Institute of Brewing, **109** (2003), no. 2, pp. 106-109.
12. Aldred, P.; Kanauchi, M. and Bamforth, C.W.: An investigation into proteolysis in mashing, Journal of the Institute of Brewing, **127** (2021), no. 1, pp. 21-26.
13. Taylor, J.R.N. and Boyd, H.K.: Free α-amino nitrogen production in sorghum beer mashing, Journal of the Science of Food and Agriculture, **37** (1986), no. 11, pp. 1109-1117.
14. Zdaniewicz, M.; Satora, P.; Pater, A. and Bogacz, S.: Low Lactic Acid-Producing Strain of Lachancea thermotolerans as a New Starter for Beer Production, Biomolecules, **10** (2020), no. 2, p. 256.
15. Djukić-Vuković, A.; Mladenović, D.; Radosavljević, M.; Kocić-Tanackov, S.; Pejin, J. and Mojović, L.: Wastes from bioethanol and beer productions as substrates for l(+) lactic acid production – A comparative study, Waste Management, **48** (2016), pp. 478-482.
16. Bellut, K.; Michel, M.; Zarnkow, M.; Hutzler, M.; Jacob, F.; De Schutter, D.; Daenen, L.; Lynch, K.; Zannini, E. and Arendt, E.: Application of Non-Saccharomyces Yeasts Isolated from Kombucha in the Production of Alcohol-Free Beer, Fermentation, **4** (2018), no. 3, p. 66.
17. Lie, S.: The EBC-Ninhydrin Method for Determination of Free Alpha Amino Nitrogen, Journal of the Institute of Brewing, **79** (1973), pp. 37-41.
18. Das, A.J.; Khawas, P.; Miyaji, T. and Deka, S.C.: HPLC and GC-MS analyses of organic acids, carbohydrates, amino acids and volatile aromatic compounds in some varieties of rice beer from northeast India, Journal of the Institute of Brewing, **120** (2014), no. 3, pp. 244-252.
19. Erbe, T. and Brückner, H.: Chromatographic determination of amino acid enantiomers in beers and raw materials used for their manufacture, Journal of Chromatography A, **881** (2000), no. 1-2, pp. 81-91.
20. Lekkas, C.; Stewart, G.G.; Hill, A.E.; Taidi, B. and Hodgson, J.: Elucidation of the Role of Nitrogenous Wort Components in Yeast Fermentation, Journal of the Institute of Brewing, **113** (2007), no. 1, pp. 3-8.
21. Weishaupt, I.; Zimmer, M.; Neubauer, P. and Schneider, J.: Model based optimization of transflection near infrared spectroscopy as a process analytical tool in a continuous flash pasteurizer, Journal of Food Science, **85** (2020), no. 7, pp. 2020-2031.
22. Zimmer, M. and Schneider, J.: Near-infrared diffuse reflectance spectroscopy for discriminating fruit and vegetable products preserved in glass containers, Croatian journal of food science and technology, **11** (2019), no. 1, pp. 104-112.
23. Menesatti, P.; Zanella, A.; D'Andrea, S.; Costa, C.; Paglia, G. and Pallottino, F.: Supervised multivariate analysis of hyper-spectral NIR images to evaluate the starch index of apples, Food and Bioprocess Technology, **2** (2009), no. 3, pp. 308-314.
24. Blanco, M. and Villarroya, I.: NIR spectroscopy: A rapid-response analytical tool, TrAC - Trends in Analytical Chemistry, **21** (2002), no. 4, pp. 240-250.
25. Morsy, N. and Sun, D.W.: Robust linear and non-linear models of NIR spectroscopy for detection and quantification of adulterants in fresh and frozen-thawed minced beef, Meat Science, **93** (2013), no. 2, pp. 292-302.
26. Malegori, C.; Nascimento Marques, E.J.; de Freitas, S.T.; Pimentel, M.F.; Pasquini, C. and Casiraghi, E.: Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms, Talanta, **165** (2017), no. November 2016, pp. 112-116.
27. Liu, Y.; Sun, X.; Zhou, J.; Zhang, H. and Yang, C.: Linear and nonlinear multivariate regressions for determination sugar content of intact Gannan navel orange by Vis-NIR diffuse reflectance spectroscopy, Mathematical and Computer Modelling, **51** (2010), no. 11-12, pp. 1438-1443.
28. Ding, H.B. and Xu, R.J.: Near-Infrared Spectroscopic Technique for Detection of Beef Hamburger Adulteration, Journal of Agricultural and Food Chemistry, **48** (2000), no. 6, pp. 2193-2198.
29. Vann, L.; Layfield, J.B. and Sheppard, J.D.: The application of near-infrared spectroscopy in beer fermentation for online monitoring of critical process parameters and their integration into a novel feedforward control strategy, Journal of the Institute of Brewing, **123** (2017), no. 3, pp. 347-360.
30. Cozzolino, D.; Parker, M.; Dambergs, R.G.; Herderich, M. and Gishen, M.: Chemometrics and Visible-Near Infrared Spectroscopic Monitoring of Red Wine Fermentation in a Pilot Scale, (2006).
31. Gonza, C.; Alvarez-garcı, N. and Gonza, I.: Determination of fatty acids in the subcutaneous fat of Iberian breed swine by near infrared spectroscopy (NIRS) with a fibre-optic probe, **65** (2003), pp. 713-719.
32. Özdemir, D.: Genetic multivariate calibration for near infrared spectroscopic determination of protein, moisture, dry mass, hardness and other residues of wheat, International Journal of Food Science and Technology, **41** (2006), no. s2, pp. 12-21.
33. Gonzalez Viejo, C.; Fuentes, S.; Torrico, D.; Howell, K. and Dunshea, F.R.: Assessment of beer quality based on foamability and chemical composition using computer vision algorithms, near infrared spectroscopy and machine learning algorithms, Journal of the Science of Food and Agriculture, **98** (2018), no. 2, pp. 618-627.
34. Gonzalez Viejo, C.; Fuentes, S.; Torrico, D.D.; Howell, K. and Dunshea, F.R.: Assessment of Beer Quality Based on a Robotic Pourer, Computer Vision, and Machine Learning Algorithms Using Commercial Beers, Journal of Food Science, **83** (2018), no. 5, pp. 1381-1388.
35. Titze, J.; Ilberg, V.; Friess, A.; Jacob, F. and Parlar, H.: Efficient and quantitative measurement of malt and wort parameters using FTIR spectroscopy, Journal of the American Society of Brewing Chemists, **67** (2009), no. 4, pp. 193-199.
36. Klein, H. and Forster, C.: Analytik Vision 2020 - Überblick über zukünftige Möglichkeiten in der Brauereianalytik, Brauindustrie, **10** (2016), pp. 45-49.
37. Wiesner, K.; Fuchs, K.; Gigler, A.M. and Pastusiak, R.: Trends in near infrared spectroscopy and multivariate data analysis from an industrial

perspective, Procedia Engineering, **87** (2014), pp. 867-870.
38. Lopes, M.; Amorim, A.; Calado, C. and Reis Costa, P.: Determination of Cell Abundances and Paralytic Shellfish Toxins in Cultures of the Dinoflagellate Gymnodinium catenatum by Fourier Transform Near Infrared Spectroscopy, Journal of Marine Science and Engineering, **6** (2018), no. 4, p. 147.
39. Mendez, K.M.; Reinke, S.N. and Broadhurst, D.I.: A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification, Metabolomics, **15** (2019), no. 12, pp. 1-15.
40. Breiman, L.: Statistical modeling: The two cultures, Statistical Science, **16** (2001), no. 3, pp. 199-215.
41. Prakash, M.; Sarin, J.K.; Rieppo, L.; Afara, I.O. and Töyräs, J.: Optimal Regression Method for Near-Infrared Spectroscopic Evaluation of Articular Cartilage, Applied Spectroscopy, **71** (2017), no. 10, pp. 2253-2262.
42. Kucheryavskiy, S.: Analysis of NIR spectroscopic data using decision trees and their ensembles, Journal of Analysis and Testing, **2** (2018), no. 3, pp. 274-289.
43. Amuah, C.L.Y.; Teye, E.; Lamptey, F.P.; Nyandey, K.; Opoku-Ansah, J. and Adueming, P.O.: Feasibility Study of the Use of Handheld NIR Spectrometer for Simultaneous Authentication and Quantification of Quality Parameters in Intact Pineapple Fruits, Journal of Spectroscopy, 2019 (2019), pp. 1-9.
44. Kaufmann, K.C.; Favero, F. de F.; de Vasconcelos, M.A.M.; Godoy, H.T.; Sampaio, K.A. and Barbin, D.F.: Portable NIR Spectrometer for Prediction of Palm Oil Acidity, Journal of Food Science, **84** (2019), no. 3, pp. 406-411.
45. Cui, C. and Fearn, T.: Comparison of partial least squares regression, least squares support vector machines, and Gaussian process regression for a near infrared calibration, Journal of Near Infrared Spectroscopy, **25** (2017), no. 1, pp. 5-14.
46. Chung, H.; Lee, H. and Jun, C.: Determination of Research Octane Number using NIR Spectral Data and Ridge Regression, **22** (2001), no. 1, pp. 1-6.
47. Vigneau, E.; Devaux, M.F.; Qannari, E.M. and Robert, P.: Principal Component Regression, Ridge Regression and Ridge Principal Component Regression in Spectroscopy Calibration, **11** (1997), no. September 1996, pp. 239-249.
48. Clavaud, M.; Roggo, Y.; Dégardin, K.; Sacré, P.; Hubert, P. and Ziemons, E.: Global regression model for moisture content determination using near-infrared spectroscopy, European Journal of Pharmaceutics and Biopharmaceutics, **119** (2017), pp. 343-352.
49. Chen, T.; Morris, J. and Martin, E.: Gaussian process regression for multivariate spectroscopic calibration, **87** (2007), pp. 59-71.
50. Schoot, M.; Kapper, C.; van Kollenburg, G.H.; Postma, G.J.; van Kessel, G.; Buydens, L.M.C. and Jansen, J.J.: Investigating the need for preprocessing of near-infrared spectroscopic data as a function of sample size, Chemometrics and Intelligent Laboratory Systems, **204** (2020), no. July, p. 104105.
51. Stordrange, L.; Libnau, F.O.; Malthe-Sørenssen, D. and Kvalheim, O.M.: Feasibility study of NIR for surveillance of a pharmaceutical process, including a study of different preprocessing techniques, Journal of Chemometrics, **16** (2002), no. 8-10, pp. 529-541.
52. Xu, L.; Zhou, Y.P.; Tang, L.J.; Wu, H.L.; Jiang, J.H.; Shen, G.L. and Yu, R.Q.: Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration, Analytica Chimica Acta, **616** (2008), no. 2, pp. 138-143.
53. Thennadil, S.N. and Martin, E.B.: Empirical preprocessing methods and their impact on NIR calibrations: A simulation study, Journal of Chemometrics, **19** (2005), no. 2, pp. 77-89.
54. Mehta, P.; Bukov, M.; Wang, C.-H.; Day, A.G.R.; Richardson, C.; Fisher, C.K. and Schwab, D.J.: A high-bias, low-variance introduction to Machine Learning for physicists, Physics Reports, **810** (2019), pp. 1-124.
55. Awad, M. and Khanna, R.: Efficient learning machines: Theories, concepts, and applications for engineers and system designers, Apress, Berkeley, CA, 2015.
56. Kong, W.; Zhang, C.; Liu, F.; Nie, P. and He, Y.: Rice Seed Cultivar Identification Using Near-Infrared Hyperspectral Imaging and Multivariate Data Analysis, Sensors, **13** (2013), no. 7, pp. 8916-8927.
57. Ringnér, M.: What is principal component analysis?, Nature Biotechnology, **26** (2008), no. 3, pp. 303-304.
58. Wefing, P.; Conradi, F.; Trilling, M.; Neubauer, P. and Schneider, J.: Approach for modelling the extract formation in a continuous conducted "β-amylase rest" as part of the production of beer mash with targeted sugar content, Biochemical Engineering Journal, 164 (2020), p. 107765.
59. Steiner, E.; Gastl, M. and Becker, T.: Protein changes during malting and brewing with focus on haze and foam formation: a review, European Food Research and Technology, **232** (2011), no. 2, pp. 191-204.
60. Jones, B.L. and Marinac, L.: The Effect of Mashing on Malt Endoproteolytic Activities, Journal of Agricultural and Food Chemistry, **50** (2002), no. 4, pp. 858-864.
61. Lou, Y.; Caruana, R. and Gehrke, J.: Intelligible models for classification and regression, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2012), pp. 150-158.
62. MacGregor, A.W.; Bazin, S.L.; Macri, L.J. and Babb, J.C.: Modelling the Contribution of Alpha-Amylase, Beta-Amylase and Limit Dextrinase to Starch Degradation During Mashing, Journal of Cereal Science, **29** (1999), no. 2, pp. 161-169.
63. Inagaki, T.; Shinoda, Y.; Miyazawa, M.; Takamura, H. and Tsuchikawa, S.: Near-infrared spectroscopic assessment of contamination level of sewage, Water Science and Technology, **61** (2010), no. 8, pp. 1957-1963.
64. Kneib, T.; Konrath, S. and Fahrmeir, L.: High dimensional structured additive regression models: Bayesian regularization, smoothing and predictive performance, Journal of the Royal Statistical Society. Series C: Applied Statistics, **60** (2011), no. 1, pp. 51-70.
65. Schaffer, C.: Overfitting Avoidance as Bias, Machine Learning, **10** (1993), no. 2, pp. 153-178.
66. Xu, W.; Liu, X.; Leng, F. and Li, W.: Blood-based multi-tissue gene expression inference with Bayesian ridge regression, Bioinformatics (Oxford, England), **36** (2020), no. 12, pp. 3788-3794.
67. Shi, Z.; Ji, W.; Viscarra Rossel, R.A.; Chen, S. and Zhou, Y.: Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library, European Journal of Soil Science, **66** (2015), no. 4, pp. 679-687.
68. Goodfellow, I.; Bengio, Y. and Courville, A.: Deep Learning, First Edit ed., MIT Press, 2016.
69. MacKay, D.J.C.: Bayesian Interpolation, Neural Computation, **4** (1992), no. 3, pp. 415-447.
70. Tipping, M.E.: Sparse Bayesian Learning and the Relevance Vector Machine, Journal of Machine Learning Research, **1** (2001), no. 3, pp. 211-244.
71. Meade, N.: A comparison of the accuracy of short term foreign exchange forecasting methods, International Journal of Forecasting, **18** (2002), no. 1, pp. 67-83.

72. Drucker, H.; Surges, C.J.C.; Kaufman, L.; Smola, A. and Vapnik, V.: Support vector regression machines, Advances in Neural Information Processing Systems, 1997, pp. 155-161.
73. Richman, M.B. and Adrianto, I.: Classification and regionalization through kernel principal component analysis, Physics and Chemistry of the Earth, **35** (2010), nos. 9-12, pp. 316-328.
74. Smola, A.J. and Schölkopf, B.: A tutorial on support vector regressionStatistics and Computing2004.
75. Schrider, D.R. and Kern, A.D.: Supervised Machine Learning for Population Genetics: A New Paradigm, Trends in Genetics, **34** (2018), no. 4, pp. 301-312.
76. Eisenberg, D.T.A.; Kuzawa, C.W. and Hayes, M.G.: Improving qPCR telomere length assays: Controlling for well position effects increases statistical power, American Journal of Human Biology, **27** (2015), no. 4, pp. 570-575.
77. Lavanya, G.; Sunil, M.; Eswarudu, M.M.; Chinna Eswaraiah, M.; Harisudha, K. and Naga Spandana, B.: Analytical Method Validation: An Updated Review, Int J Pharm Sci Res, **4** (2013), no. 4, pp. 1280-1286.
78. ICH: Harmonised Tripartie Guide Prepared within the Third Inter- national Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH), Validation of analytical Procedures: Methodology, (1996), pp. 1-8.
79. Singla, M.; Ghosh, D. and Shukla, K.K.: A survey of robust optimization based machine learning with special reference to support vector machines, International Journal of Machine Learning and Cybernetics, **11** (2020), no. 7, pp. 1359-1385.
80. Hancox-Li, L.: Robustness in machine learning explanations, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, ACM, New York, NY, USA, 2020, pp. 640-647.
81. Sehwag, V.; Bhagoji, A.N.; Song, L.; Sitawarin, C.; Cullina, D.; Chiang, M. and Mittal, P.: Analyzing the Robustness of Open-World Machine Learning, Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security - AISec'19, ACM Press, New York, New York, USA, 2019, pp. 105-116.
82. Efendi, A. and Effrihan: A simulation study on Bayesian Ridge regression models for several collinearity levels, AIP Conference Proceedings, 1913 (2017), no. December 2017.
83. Ferragina, A.; de los Campos, G.; Vazquez, A.I.; Cecchinato, A. and Bittante, G.: Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data, Journal of Dairy Science, **98** (2015), no. 11, pp. 8133-8151.
84. Dietterich, T.: Overfitting and undercomputing in machine learning, ACM Computing Surveys, **27** (1995), no. 3, pp. 326-327.
85. Marc, A.; Engasser, J.M.; Moll, M. and Flayeux, R.: A kinetic model of starch hydrolysis by alpha- and beta-amylase during mashing., Biotechnology and bioengineering, **25** (1983), no. 2, pp. 481-96.
86. Wijngaard, H.H. and Arendt, E.K.: Optimisation of a Mashing Program for 100% Malted Buckwheat, Journal of the Institute of Brewing, **112** (2006), no. 1, pp. 57-65.
87. Kettunen, A.; Hamalainen, J.J.; Stenholm, K. and Pietila, K.: A model for the prediction of beta-glucanase activity and beta-glucan concentration during mashing, Journal of Food Engineering, **29** (1996), pp. 185-200.
88. Robinson, L.; Evans, D.; Kaukovirta-Norja, A.; Vilpola, A.; Aldred, P. and Home, S.: The interaction between malt protein quality and brewing conditions and their impact on beer colloidal stability, MBAA Technical Quarterly, **41** (2004), no. 4, pp. 353-362.
89. García, S.; Ramírez-Gallego, S.; Luengo, J.; Benítez, J.M. and Herrera, F.: Big data preprocessing: methods and prospects, Big Data Analytics, **1** (2016), no. 1, pp. 1-22.
90. Cao, X.H.; Stojkovic, I. and Obradovic, Z.: A robust data scaling algorithm to improve classification accuracies in biomedical data, BMC Bioinformatics, **17** (2016), no. 1, p. 359.
91. Abdi, H. and Williams, L.J.: Principal component analysis, Wiley Interdisciplinary Reviews: Computational Statistics, **2** (2010), no. 4, pp. 433-459.
92. Lever, J.; Krzywinski, M. and Altman, N.: Principal component analysis, Nature Methods, **14** (2017), no. 7, pp. 641-642.

Received 29 June 2021, accepted 13 September 2021