

PREDICTION OF PROXIMATE ANALYSIS AND PROCESS TEMPERATURE OF TORREFIED AND PYROLYZED WOOD PELLETS BY NEAR-INFRARED SPECTROSCOPY COUPLED WITH MACHINE LEARNING

Meltem Kasapoglu-Calik Department of Chemical Technology, Istanbul Gedik University, Turkey, meltem.calik@gedik.edu.tr

([ID](https://orcid.org/0000-0001-5851-0714)) <https://orcid.org/0000-0001-5851-0714>)

*Ebubekir Siddik Aydin, Department of Chemical Engineering, Gebze Technical University, Turkey, e.aydin@gtu.edu.tr

([ID](https://orcid.org/0000-0002-8704-4502)) <https://orcid.org/0000-0002-8704-4502>)

Ozgun Yucel, Department of Chemical Engineering, Gebze Technical University, Turkey, yozgun@gtu.edu.tr

([ID](https://orcid.org/0000-0001-8916-2628)) <https://orcid.org/0000-0001-8916-2628>)

Received: 27.08.2020, Accepted: 14.12.2020

Research Article

*Corresponding author

DOI: 10.22531/muglajsci.785974

Abstract

Near-Infrared (NIR) Spectroscopy is a time and cost-effective method to characterize the materials in the food, petrochemical, pharmaceutical, and agricultural industries. Proximate analysis of the carbon-containing materials and investigating the effectiveness of the heat treatments on the material are a particularly time-consuming process. This work presents the four regression methods, i.e., decision tree regression, support vector regression and two versions of ensembles of decision trees to predict the proximate analysis of biomass and heat treatment temperature. Thus, effective method has been proposed to reduce experimental effort and present the characterization of heat-treated biomass feedstock theoretically. Prediction results show that SVR and ENS2 regression methods calibrating the NIR spectra to the values of wood pellet properties achieved good performance with the coefficient of determination (R^2) of 0.880- 0.984 and RMSE of 0.444- 5.308 for ash and volatile matter. This study suggests that machine learning-based regression methods with integrated NIR spectroscopy of biomass is promising as an alternative method for rapid characterization. Another possible application of the current study is that it can be used for processed fuel recognition prior to a fully automated fuel quality assessment system in the biomass industry.

Keywords: biomass, near-infrared spectroscopy, machine learning, support vector regression, ensembles of decision trees, decision tree regression

TORREFIYE VE PİROLİZ EDİLMİŞ ODUN PELETİNİN PROSES SICAKLIĞININ, KISA ANALİZİNİN MAKİNE ÖĞRENMESİ DESTEKLİ YAKIN KIZILÖTESİ SPEKTROSKOPİSİ İLE TAHMİNİ

Özet

Yakın Kızılötesi (NIR) Spektroskopisi, gıda, petrokimya, ilaç ve tarım endüstrilerindeki malzemeleri karakterize etmek için zaman ve maliyet açısından etkin bir yöntemdir. Karbon içeren malzemelerin yakın analizi ve ısı işlemlerinin malzeme üzerindeki etkinliğinin araştırılması özellikle zaman alan bir süreçtir. Bu çalışmada, biyokütle ve ısı işlem sıcaklığının yakın analizini tahmin etmek için dört regresyon yöntemi olan karar ağacı regresyonu, destek vektör regresyonu ve rassal orman regresyonunun iki versiyonu kullanılmıştır. Deneysel çabayı azaltmak ve ısı işlem görmüş biyokütle hammaddesinin karakterizasyonunu teorik olarak sunmak için etkili bir yöntem önerilmiştir. Tahmin sonuçları, odun peletinin NIR spektrum değerlerini kalibre eden SVR ve ENS2 regresyon yöntemlerinin kül ve uçucu madde için 0.880- 0.984 belirleme katsayısı (R^2) ve 0.444- 5.308 RMSE ile iyi performans elde ettiğini göstermektedir. Bu çalışma, entegre NIR spektroskopisine sahip makine öğrenmesine dayalı regresyon yöntemlerinin biyokütlenin hızlı karakterizasyon için alternatif bir yöntem olarak umut verici olduğunu göstermektedir. Mevcut çalışmanın bir başka olası uygulaması, biyokütle endüstrisinde tam otomatik yakıt kalitesi değerlendirme sisteminden önce işlenmiş yakıt tanıma sistemleri için kullanılabilir.

Anahtar Kelimeler: biyokütle, yakın kızılötesi spektroskopisi, makine öğrenimi, destek vektör regresyonu, rassal orman, karar ağacı

Cite

Kasapoglu-Calik, M., Aydin, E. S. and Yucel, O., (2020). "Prediction of proximate analysis and process temperature of torrefied and pyrolyzed wood pellets by near-infrared spectroscopy coupled with machine learning", *Mugla Journal of Science and Technology*, 6(2), 99-110.

1. Introduction

According to the recent Global Energy and CO₂ Status Report of the International Energy Agency (IEA), in 2018, the energy-related carbon dioxide (CO₂) emissions grew by 1.7% to 33.1 billion tons. It was the highest growth rate estimated since 2013 and was 70% higher than the average increase since 2010 [1]. The increase in the CO₂ emissions will cause irreversible climatic, environmental and ecological changes of the Earth's atmospheric system on both short and long-time scales [2]. By year of 2019, about 81.4% of the world's energy demand are provided from three main fossil fuels, oil, coal and natural gas [3]. The continued use of fossil fuels to generate energy has the potential to further intensify global CO₂ emissions and climate change, as well as negative social, political and environmental impacts [3, 4]. Therefore, it is necessary to establish a global, sustainable and bio-based economy to reduce greenhouse gas emissions and increase climate resistance [5].

Among the various renewable energies, biomass energy conversions have seen large-scale implementations worldwide. The world energy consumption based on biomass has increased to 14% of total energy consumption [6]. Biomass energy conversion is a reliable and cost-effective method to produce chemicals and meet energy demand without requiring storage compared to other renewable sources such as wind, solar and wave [7]. Biomass is expected to have the potential to become one of the major global energy sources in the next century. It is thought that modernized bioenergy systems will contribute to sustainable energy systems in the future and sustainable development in industrialized as well as developing countries [8]. Biomass can be used to generate energy with various technologies, but when raw biomass is used as fuel, it has some disadvantages such as low calorific value, high moisture content, hygroscopic nature [9]. In order to reduce these disadvantages, biomass is treated with thermal energy yielding processes such as torrefaction, pyrolysis and gasification, which produces biofuels. Biomass torrefaction is a mild pretreatment of biomass carried out at temperatures of 200 – 340 °C. Since most of the moisture, volatiles and hemicellulose are removed from the biomass in the torrefaction process, the biomass loses its tenacious and fibrous structure. Eventually, the resulting torrefied biomass has higher energy density, and improved ignitability, reactivity and grindability with more uniform structure when compared to raw biomass [10]. Given these advantages, the torrefied biomass can be regarded as a more valuable fuel than raw biomass.

Biochar is a carbon-rich product obtained when biomass is heated at temperatures below 700 °C in a closed container with limited or no available oxygen. In addition to being an excellent material for energy production, biochar has also been described as an advantageous material to improve soil fertility and

mitigate climate change by reducing CO₂ emissions [11]. Although the quality of the biochar depends mainly on the type of raw material and the pyrolysis conditions, determining the quality and properties of the biochar is costly and time-consuming process [12]. As a result, alternative techniques are needed to effectively determine parameters of torrefaction, pyrolysis and gasification processes.

Near Infrared Spectroscopy is a time and cost-effective method commonly used in the food, petrochemical, pharmaceutical and agricultural industries [13-16]. The NIR technology provides information with sufficient precision for solid and liquid systems without any sample preparation. This technology has also been increasingly used in the wood industry to determine physical, mechanical properties and the amount of wood components such as lignin, cellulose, hemicellulose, extract, etc. [17, 18]. So et al. determined that the rapid assessment of solid wood properties using NIR spectra. They also noted that NIR as an online monitoring tool in the manufacturing process encourages many laboratories to examine potential applications for wood composites [19]. In addition to predicting wood properties, NIR spectroscopy was also used to classify wood species and their origins [20, 21]. Schwanninger et al. reported that there is a close relationship between the chemical changes of thermally modified wood and its corresponding NIR spectra [22]. NIR radiation interacts with polar structural groups such as, C-H, C-O, C-O-H, O-H, N-H and C=O bonds. The absorption bands observed in NIR spectra of biomass arise from overtones and combinations of these bonds. Due to the overtone vibrations in the NIR spectra provide valuable chemical information about the state of torrefied biomass, NIR techniques can be used to monitor and control biomass treatment and conversion processes as well as the resulting torrefaction quality. The characteristics of treated biomass are highly related to chemical modifications induced by temperature conditions and treatment duration [23]. So far, some studies have been reported for using NIR spectroscopy to identify the raw wood components such as lignin, cellulose, hemicellulose, etc. [24, 25]. Multivariate modeling of the resulting NIR data can be used to accurately estimate not only the properties of the torrefied wood such as volatile matter, fixed carbon and ash content, but also important process parameters such as temperature and time. Via et al. built calibration models of the NIR data for the prediction of proximate analysis after torrefaction [26]. However, there is a few researches on identification of process parameters of torrefied biomass, biochar and the solid residue after gasifying biomass. Rousset et al. applied principal component analysis to NIR spectra of thermally treated wood samples. The authors indicated that NIR spectroscopy can effectively adapted to the rapid measurement of native and thermally treated wood [9].

As NIR spectroscopy is rapid and completely non-destructive technique it would be an attractive option for utilized in standardized procedures for measuring diverse variables. Therefore, it would be possible to develop NIR-based methods which have major advantages over traditional analysis techniques for characterizing torrefied biomass, biochar and solid residues after gasification. Real-time monitoring of the process using the NIR spectroscopy technique may be an advantageous method to ensure the quality of the final product to the consumer.

In this study, the effect of process temperature and heat treatment time on torrefied and pyrolyzed wood pellets were studied. After the heat treatment process, the proximate analysis of torrefied/pyrolyzed biomass was also studied. The NIR spectra of torrefied and pyrolyzed biomass at different heating temperature and duration were collected. In the last part of the study, four regression methods including support vector regression (SVR), decision tree regression (DTR), and the ensemble of decision trees were utilized to build prediction models for the above-mentioned outputs.

2. Methods

2.1. Heat Treatment

The wood chips in pellet form were kindly provided from a local market in Ankara, TURKEY. Biochar samples were obtained by torrefaction and pyrolysis of raw biomass at various temperatures, allowing the materials to carbonize instead of combustion under oxygen-limited conditions. The heat treatment procedure was performed in a tubular furnace at the temperatures of 200, 250, 300, 350, 400 and 450 °C with a residence time of 0.25, 0.50, 0.75, 1, 2, 3, 4, 5 and 6 hours under continuous nitrogen flow (150 ml/min). A total of 54 torrefied and pyrolyzed samples were collected. The samples were kept in a desiccator until they were used. The proximate analyses of wood pellets were determined using standard procedures. ASTM E1755-01 for ash, ASTM E871-72 for moisture content and ASTM D3175 for volatile matter methods were applied for proximate analysis by using ash furnace. Fixed carbon was determined by balance.

2.2. NIR Spectroscopy

Diffuse-reflectance spectra were collected using a Fourier transform near infrared (FT-NIR) spectrophotometer (Bruker Optics, Ettlingen, Germany) and the Bruker software OPUS 5.5. The NIR spectra were acquired by an integrating sphere scanning an area of about 1 cm in diameter; 32 scans were averaged per scan at a spectral resolution of 8 cm⁻¹ between 12500 and 4000 cm⁻¹. Measurements were first taken on the gold reference (32 readings), followed by 32 readings on the sample. NIR data points were collected from a sample in log of inverse reflectance log₁₀(1/R) form (1 being the gold reference and R the reflectance of the sample for each wavelength), i.e. equivalent to an absorbance

spectrum. Three spectra were obtained for each sample, and the average of the three was recorded.

2.3. Machine Learning

2.3.1. Feature Extraction and Preprocessing

The NIR spectra may show systematic variations such as random noise and basal deviation. The pretreatment of the spectrum can reduce the complexity of the model and improve interpretation by making it more robust and reliable against undesirable variations. In this study, no pre-treatment method that comes with application software of NIR spectroscopy was used. But, we employ relative second order derivative which is given in Eq. (1) in order to eliminate exponential trends which, for example, may be caused by inflationary influences.

$$\tilde{x}[i] = \frac{x[i-1] - 2x[i] + x[i+1]}{x[i]} \quad (1)$$

where $x[i]$ is the i^{th} element of the spectrum to construct scaled features $\tilde{x}[i]$.

2.4. Support Vector Regression (SVR)

The support vector machine (SVM) is a supervised learning algorithm widely used for data classification introduced by Vapnik and Cortes [27]. It ensures high generalization ability through non-linear mapping of input vectors by constructing a decision surface [28, 29]. SVMs have been effectively used to perform not only binary but also multi-class classification and regression problems [30]. Compared to other proposed methods such as polynomial regression SVMs offer some primary advantages. It can learn in high-dimensional characteristic space by using small number of training samples [31] and over-fitting in the input space which caused by the classification in high dimension feature spaces can be controlled easily by using SVMs [32]. When the problem has complex and non-linear dynamics, the use of kernel function gives SVMs flexibility by implicitly converting the data to a higher-dimensional feature space. By using a kernel function, linearly inseparable input data becomes separable. Support vector regression (SVR) is one of the fundamental class of SVMs which has been effectively utilized for nonlinear frameworks [31]. The global prediction function of SVR is presented in Eq. (2). The kernel function defines the features space, where data are regressed. Hence, selecting an appropriate one during the regression is essential. In this study linear function Eq. (3) is used as kernel.

$$f(x) = \sum_{i=1}^n \alpha_i * K(x_i, x_j) + b \quad (2)$$

$$K(x_i, x_j) = x'_i x_j \quad (3)$$

where $\alpha_i (i = 1, 2, 3, \dots, n)$ are the support vectors, b is the bias term that is estimated in training phase and $K(x_i, x_j)$ is the kernel function. The empirical risk minimization approach (J , Eq. (4)) with robust

ϵ –insensitive loss function (L_ϵ , Eq. (5)) is used to train SVR model. α vector is obtained via the minimization of J and the constant b are used to provide estimations using Eq. (2).

$$J = \sum_{i=1}^m L_\epsilon(f(x_i), y) \quad (4)$$

$$L_\epsilon = \begin{cases} 0, & \text{if } |y-f(x)| \leq \epsilon \\ |y-f(x)| - \epsilon, & \text{otherwise} \end{cases} \quad (5)$$

here y states the output vector and ϵ defines epsilon value.

2.4.1. Decision Tree Regression (DTR)

Decision tree is widely used and advantageous machine learning method that quick to build and easy to interpret. The estimations based on decision trees are effective [33]. The decision trees are built through repeatedly dividing the data into binary sections. In each iteration the data is split according to the values of a selected attribute. The selection repeats until ‘pure’ data subsets which only include instances of the same class or pre-specified number of nodes is reached [34]. Concisely, the basic concept of a decision tree is to separate a complex decision into simpler decisions which made estimation easier to predict [35]. If predictable target attribute consists of discrete data, the developed decision tree model is called as a classification tree. If the target attribute is a continuous variable, then the model is called as a regression tree [36]. When this algorithm is applied to regression problems, it is called DTR. DTR utilizes the sum of square error metric to build subsets which has the least reasonable variations among the elements. Next, the questions in the decision tree can be followed again to obtain a continuous variable with a set of properties for the prediction.

2.4.2. Ensembles of Decision Trees: Bagging and Boosting

The main idea behind the ensemble modeling is to build a predictive model by integrating multiple models in order to obtain a classifier that outperforms every one of them. Ensembles of decision tree method can be defined as a supervised learning algorithm which combines several decision trees instead of using a single decision tree for improving prediction performance. Two of the most popular techniques for constructing ensembles are bagging and boosting [37, 38]. Bagging is a general technique for combining the predictions of many models. In bagging, the data from training samples is chosen randomly in order to create several subsets by replacing the data. In boosting, the weak learners converted to strong ones by repeatedly running a weak learner on various distributed training data. The classifiers first produced by the weak learners, then they combined into a single composite strong classifier [39]. In this study, ensembles of decision tree (bagging) is named as ENS1

and ensembles of decision tree (boosting) is named as ENS2.

3. Results and Discussion

3.1. Spectroscopic Characterization

NIR absorbance spectra of torrefied and pyrolyzed biomass samples at the temperatures of 200, 250, 300, 350, 400 and 450 °C for the heating times of 0.25, 0.5, 0.75, 1 - 6 hours are shown in Figure 1a. Since only small chemical information could be obtained from the raw spectrum, the second derivative spectra were also taken for better interpretation (Figure 1b). The differences between the absorbance spectrum of control and the heat-treated (0.25, 2, 4 and 6 h) samples are shown in Figure 2. The spectrum of untreated control sample showed a characteristic peak at 8250 cm^{-1} corresponding to the 2nd overtone of C–H stretching vibration of cellulose molecule [40]. The peaks at 6913, 6727 and 5191 cm^{-1} were attributed to the 1st overtone of O–H stretching vibrations of lignin, cellulose and including water molecules, respectively [40, 41]. The peaks at 5976 cm^{-1} and 5800 cm^{-1} corresponded to the aromatic C–H stretching vibrations of lignin and 1st overtone of C–H stretching vibration of hemicellulose, respectively [42, 43]. The peak at 4280 cm^{-1} was assigned C–H stretchings and deformations of cellulose and hemicellulose. The peak at 4411 cm^{-1} corresponded to the O–H and C–H stretching vibrations of lignin [44]. The spectral profile of 0.25 h heat-treated samples were found similar to that of the control sample. Same peaks were observed but the intensities of the peaks were displayed to be weaker compared with control. In other words, at 0.25 h duration at 200 °C, there was not much depolymerization of cellulose, hemicellulose and lignin. In the case of 2,4 and 6 h treatments, the differences were more obvious. The peaks at 6727 and 5051 cm^{-1} which corresponded to the O–H stretching vibrations of cellulose and water were disappeared after 2 h heat treatment. It was observed for the 4 h heat-treated samples that the broad band around 5600 cm^{-1} disappeared which assigned to the C–H stretching vibrations of semi-crystalline and crystalline regions in cellulose [42]. The peak at 8250 cm^{-1} which belongs to the C–H stretching vibrations of cellulose molecule were also disappeared. The intensity of the peak at 6727 cm^{-1} which attributed to the cellulose decreased after the 4 h heat treatment. At the end of the 6 h heat treatment, the mentioned peak disappeared. The disappearance of the peaks can be attributed to the decomposition of intramolecular H bonding interactions and C–H groups which associated with cellulose and hemicellulose molecules [44]. In the case of 6 h heat treatment, it was observed that the peak at 4280 cm^{-1} which belongs to the cellulose and hemicellulose molecule vibrations were also completely disappeared.

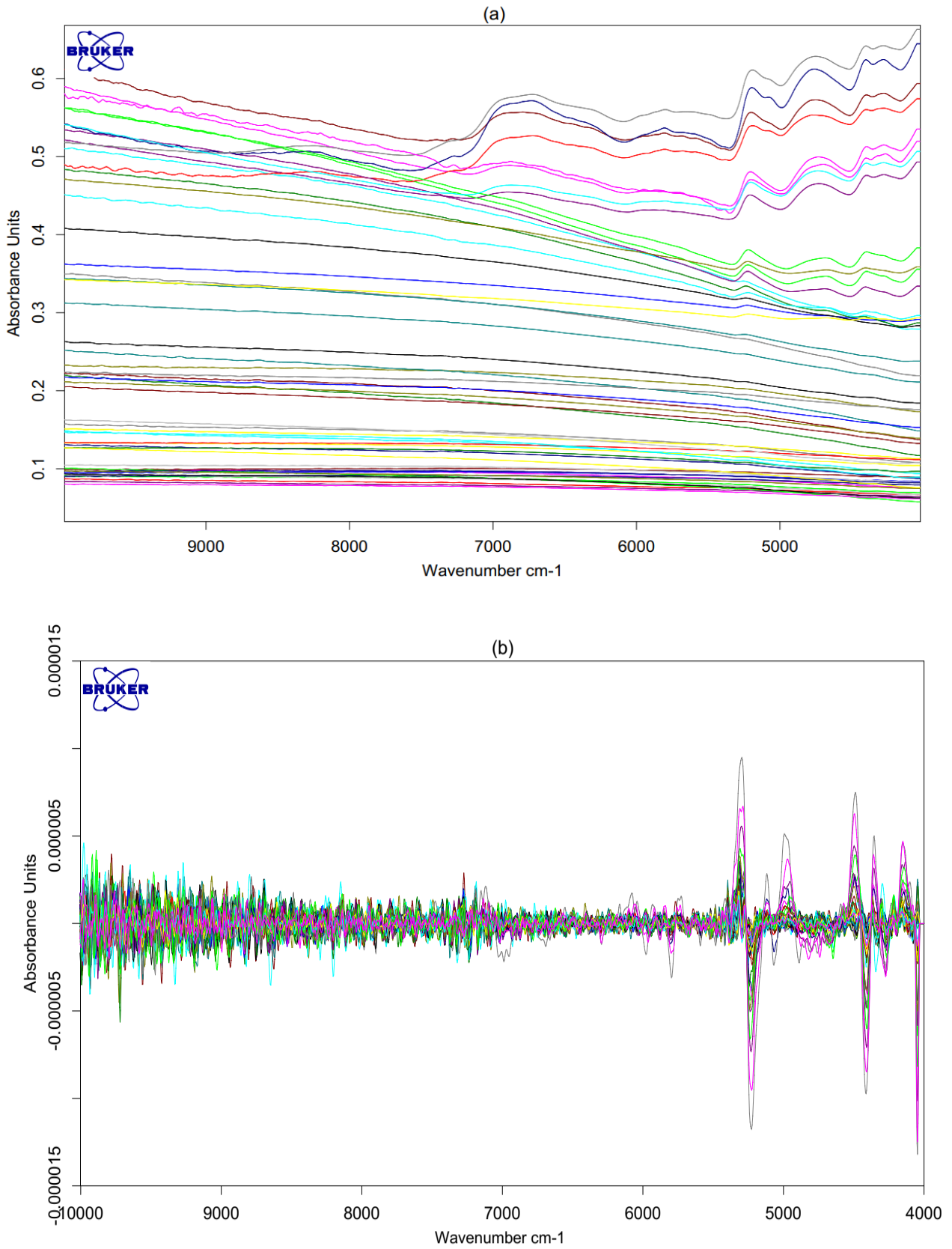


Figure 1. NIR absorbance spectra (a) and 2nd derivative spectra (b) of 54 torrefied and pyrolyzed biomass samples.

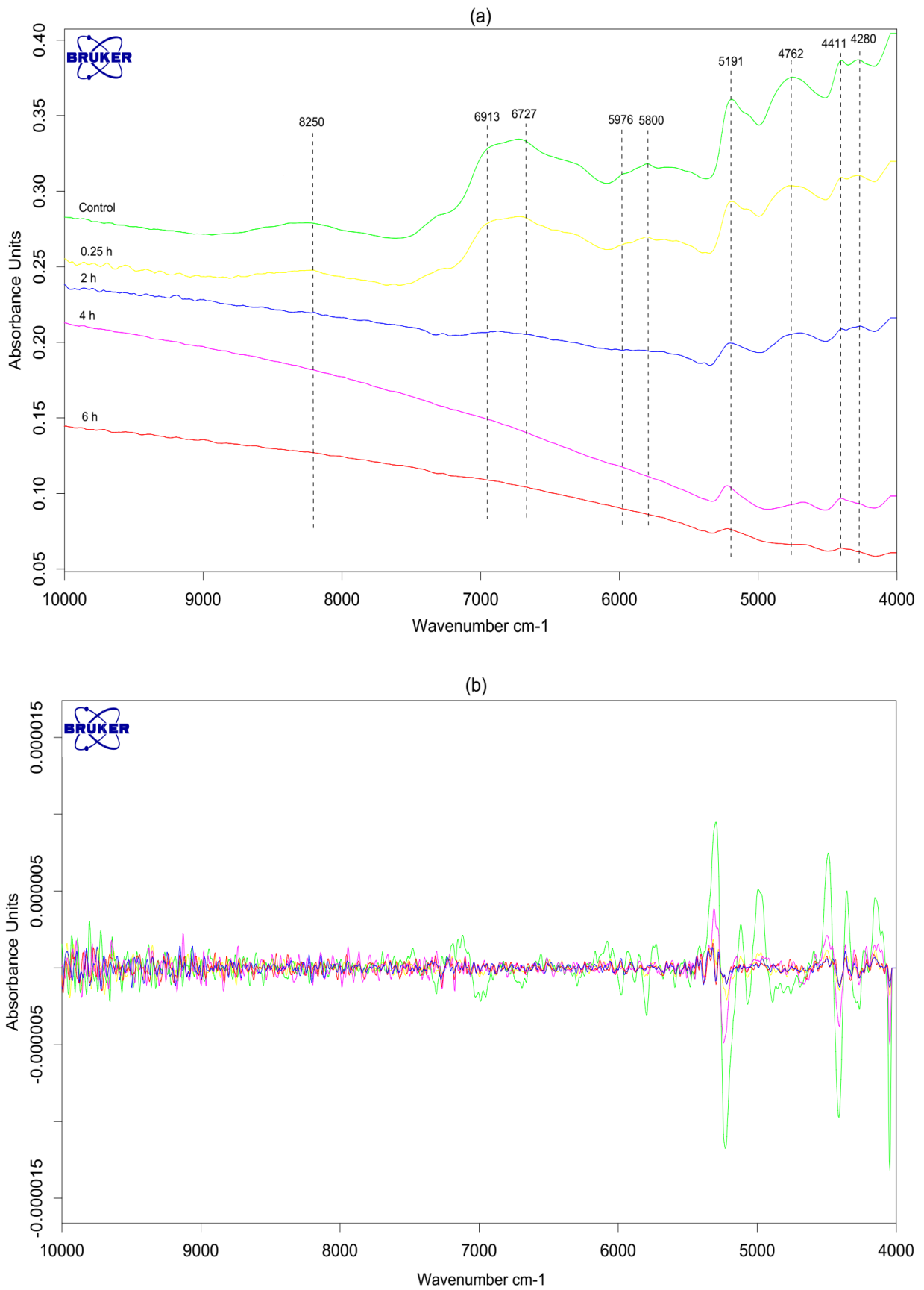


Figure 2. NIR absorbance spectra (a) and 2nd derivative spectra (b) of control, 0.25, 2, 4 and 6 h heat-treated samples.

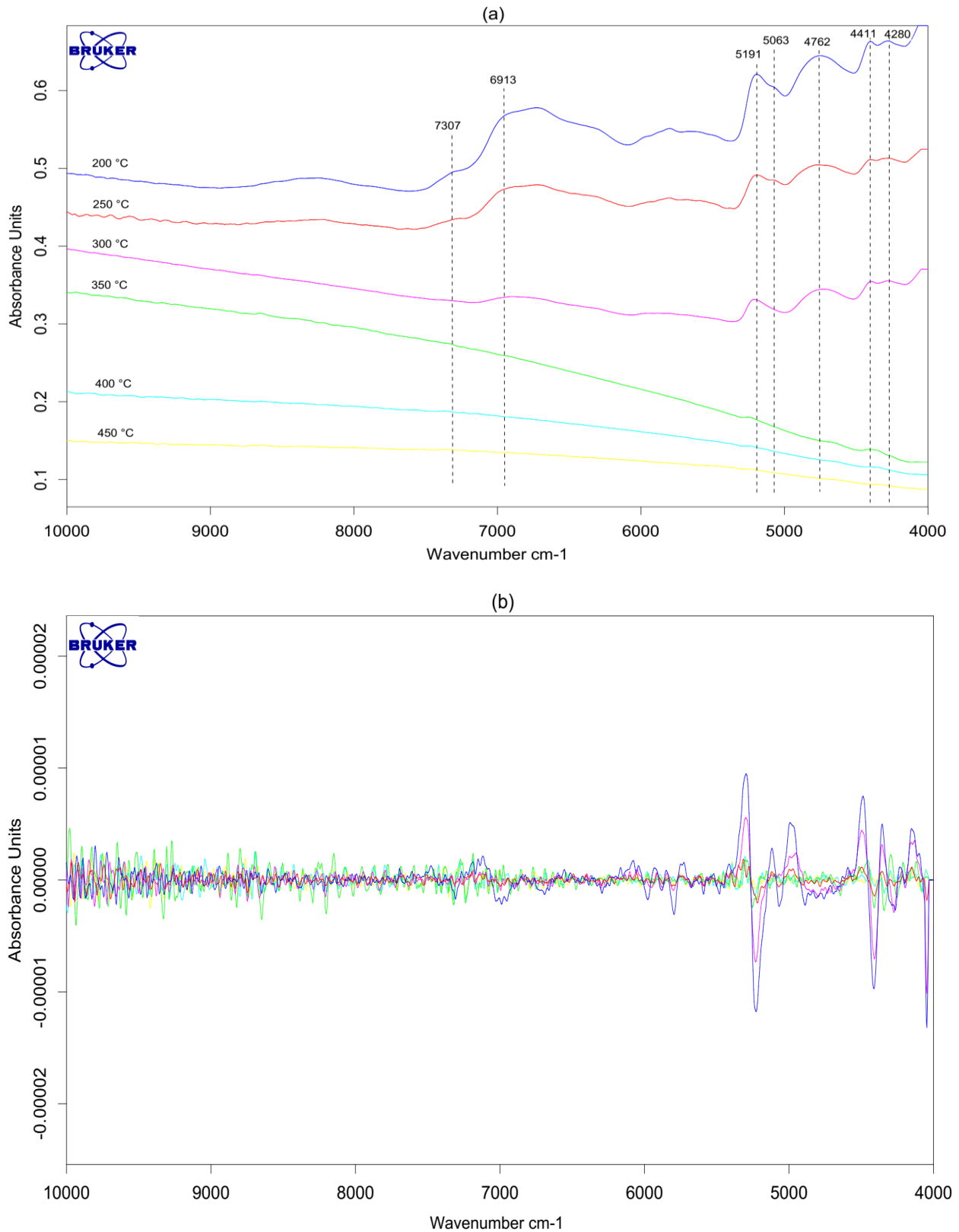


Figure 3. NIR absorbance spectra (a) and 2nd derivative spectra (b) of heat-treated samples at 200, 250, 300, 350, 400 and 450 °C.

The intensity of the peak at 5191 cm⁻¹ which belongs to the O-H groups in water has shown to be decreased as the heat treatment time increased, due to the reduction of O-H groups. The NIR spectra of heat-treated samples between the temperatures of 200 - 450 °C are shown in Figure 3. In the spectra at the temperatures between 200 - 300 °C, a downward shift of the baseline was observed which can be attributed to the degradation of hemicellulose. The peak at 7307 cm⁻¹ which belongs to the C-H stretching of hemicellulose molecule disappeared as the temperature increased to 300 °C. It is known that hemicelluloses are the most unstable components of wood with respect to temperature [9]. This result is in accordance with the results obtained Via et al. who torrefied the wood biomass samples at 225 °C [26]. At 300 °C, the peak at 5063 cm⁻¹ corresponding to the stretching and deformation vibrations of O-H at water was disappeared due to the evaporation of water [45]. At the same temperature, there was not much depolymerization of some components at the wavenumbers of 5191, 4762 and 4400 cm⁻¹ was observed. These peaks can be attributed to the O-H stretching and deformation vibration of bound water, O-H and C-H deformation stretching of cellulose and C-H₂ stretching and deformation of cellulose, respectively [46-48]. However, it was noticed that the higher temperatures such as 350 - 400 °C caused depolymerization. Melkior et al. reported that depolymerization of wood begins at around 200 °C and increases as the temperature was increased. It was also concluded by the researchers that the cellulose was completely decomposed after treatment of wood at 350 °C [49]. The peaks at 6913 and 4411 cm⁻¹ which can be associated with vibrations of lignin molecule appeared more resistant to the temperature increase. Lignin molecule was found resistant to the thermal decomposition of the polymer chains during torrefaction process, conversely to cellulose and hemicellulose molecules. These findings are consistent with the results reported Via et al. who also reported that the lignin-based polymers were resistant to thermal depolymerization during torrefaction [26]. Based on the NIR results it could be suggested that the temperature and heat treatment time clearly affect the chemical composition of the biomass and this effect can be measured by the NIR technique.

3.2. Machine Learning

The four regression methods presented were used to improve the prediction capabilities of models for five different outputs using 2203 NIR spectra points generated by NIR absorbance spectra of torrefied and pyrolyzed biomass samples as mentioned in Section 3.1. In order to predict each output variable, all prediction models were trained and interpreted with respect to their test performance in hold-out cross-validation (absorbance data were randomly split into two parts: 15% for testing, and 85% for training) by

using correlation coefficient (R^2), root mean square error (RMSE) and normalized root mean square error (NRMSE) metrics.

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}} \quad (7)$$

$$NRMSE = \frac{1}{m} \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\bar{y}}} \quad (8)$$

where m is the number of features in the data set, \bar{y} implies the mean of output vector and \hat{y} defines the estimation vector. The efficiency of the suggested methods in estimating the five different outputs variables is given in Table 1. The estimated outputs of all samples against actual observation plots are illustrated in Figure 4 to clearly present the performance of the regression models. In the ideal case scenario, all prediction and observation points equal to each other and the slope is 1. As can be seen in Table 1, SVR and ENS2 regression methods provided the highest performance results for all output variables among other methods. In these results, R^2 is higher and RMSE and NRMSE are lower. For all outputs, ENS2 prominently outperformed SVR and other regression methods. When the characterization of complex bond interactions by NIR Spectroscopy is considered, ENS2 is generally assumed to show better performance confronted with other machine learning methods due to its ability to combine several models with lower errors by reducing pitfalls of the single model. On the other hand, SVR showed parallel prediction efficiency compared to ENS2 for entire outputs. SVR is the extension of a classification algorithm (SVM) and this is one of the reasons for the superior performance. When Figure 4a is examined, predictions by ENS2 tend to scatter on the 45-degree line but also show irrelevant scattering at some points. This is caused by strict relation between temperature and time effects on the pyrolysis. For example, long periods at low temperature may have the same effect on the biomass as short periods at high temperature. On the contrary, as can be seen on Figure 4b, predictions of process temperature improved significantly compared to prediction of process time. It inclines to form a group of output variables during the training phase. Thanks to this behavior, it can be helpful when samples of a particular output variable are within a certain range and entire estimates fall within a specific area as shown in Figure 4a [50]. For ash and volatile matter content of processed biomass, ENS2 were able to achieve satisfactory performance. However, the results of the fixed carbon predictions were satisfactory only for those greater than 32 (shown as dashed vertical line), as can be seen in Figure 4d. This behavior shows that fixed carbon prediction of processed biomass much easier compared

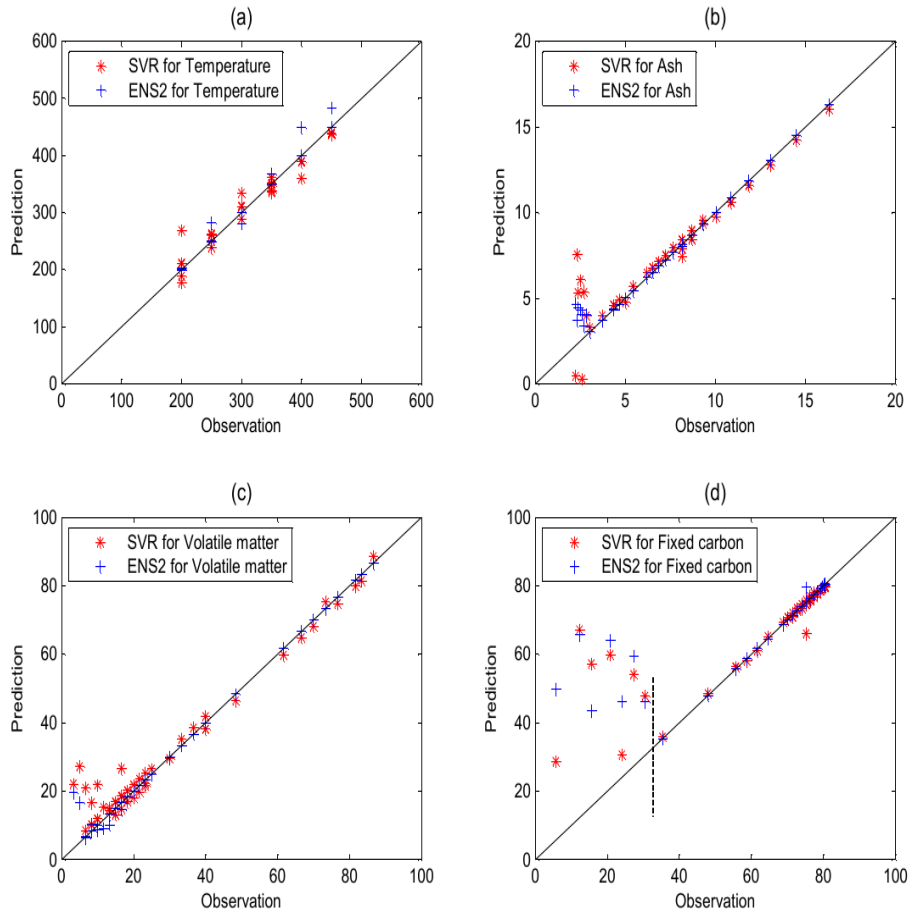


Figure 4: Predictions obtained by all proposed regression methods against observations.

to unprocessed biomass. Because as the process temperature and time increases, the amount of fixed carbon increases. However, DTR has less performance for all outputs compared to SVR and overall performance is worse than ENS2. This machine learning method is used to create prediction models based on basic logical expressions, and since it uses logic instead of numbers, it can create problems in practical processes such as characterization of biomass. For better results, it is possible to increase the success rate in processes involving more data samples [51]. In order to further evaluate the performance of the machine learning methods in the characterization of torrefied and pyrolyzed pellets more deeply, the estimation capability of each output variable was investigated. For temperature output, ENS2 and SVR succeed a convincing estimation precision with $R^2=0.986$ and $R^2=0.965$, respectively. DTR regression method also presented higher prediction accuracy ($R^2=0.948$) but ENS1 showed less performance with $R^2=0.901$ from other machine learning methods. This prediction performance can also be seen for volatile matter and ash outputs where the proposed models had good performance values. ENS is employed in two forms, i.e., bagging and boosting in this study. The prediction performance of Bagging (ENS1)

was relatively worse compared to other methods according to Table 1 except in estimating fixed carbon value. However, the bagging version of ensembles of decision trees and other machine learning methods are not sufficient to present fixed carbon output and more experiments and NIR absorbance spectra are required. To increase the fixed carbon prediction performance, the number of samples in the data set can be increased with additional characterization. This attempt can meaningfully develop prediction efficiency due to the data hunger of machine learning methods, particularly the ENS and DTR [52].

4. Conclusion

In this study, the effectiveness of machine learning methods was investigated by characterizing heat-treated wood pellets using NIR spectroscopy. Four regression methods, DTR, SVR, ENS1 and ENS2 were proposed to predict the proximate analysis variables of biomass and heat treatment duration and temperature. In order to build estimation models, a total of 2203 features are provided from the NIR absorbance spectra of torrefied and pyrolyzed wood pellets.

Table 1. Performance parameters for the presented regression methods

Output	Model	R ²	RMSE	NRMSE
Temperature	DTR	0.948	19.499	0.061
	SVR	0.965	16.053	0.056
	ENS1	0.901	26.886	0.086
	ENS2	0.986	9.925	0.029
Ash	DTR	0.900	1.031	0.333
	SVR	0.880	1.132	0.456
	ENS1	0.824	1.369	0.402
	ENS2	0.967	0.598	0.164
Volatile Matter	DTR	0.885	7.556	0.434
	SVR	0.943	5.308	1.047
	ENS1	0.900	7.018	0.929
	ENS2	0.984	2.803	0.652
Fixed Carbon	DTR	0.530	13.637	1.336
	SVR	0.627	12.138	0.935
	ENS1	0.510	13.922	1.382
	ENS2	0.567	13.086	1.256

ENS2 and SVR are carried out particularly preferable than the rest of the methods by accomplishing $R^2 > 0.88$ for temperature, ash, and volatile matter outputs and $R^2 > 0.62$ for fixed carbon output. Generally, the results revealed that NIR Spectroscopy coupled with machine learning as an alternative method for rapid characterization. The efficiency of proposed regression methods decreased while estimating fixed carbon value compared to other outputs, thus, there is a gap for progress for both feature selection and hyper-parameter to develop estimation performance for this output. With the estimation model suggested in this study, we believe that investigators can obtain a fast determination methodology to predict the proximate analysis of processed biomass. The other potential application of the present study is the actual need for processed fuel recognition prior to a fully automated fuel quality assessment system in the biomass industry. In our further study, more samples will be collected to correct and optimize the model results.

5. References

1. (IEA), I.E.A., "Market Report Series: Renewables 2018", *Analysis and Forecasts to 2023*, Paris, France, 2018.
2. Aghaalikhani, A., et al., "Detailed modelling of biomass steam gasification in a dual fluidized bed gasifier with temperature variation", *Renewable Energy*, 143, 703-718, 2019.
3. Ali, M., et al., "Spectroscopic studies of the ageing of cellulosic paper", *Polymer*, 42(7), 2893-2900, 2001.
4. Aliano-Gonzalez, M.J., et al., "A screening method based on Visible-NIR spectroscopy for the identification and quantification of different adulterants in high-quality honey", *Talanta*, 203, 235-241, 2019.
5. Almeida, G., Brito, J.O., Perré, P., "Alterations in energy properties of eucalyptus wood and bark subjected to torrefaction: the potential of mass loss as a synthetic

indicator", *Bioresource technology*, 101(24), 9778-9784, 2010.

6. Alves, A., et al., "Calibration of NIR to assess lignin composition (H/G ratio) in maritime pine wood using analytical pyrolysis as the reference method", *Holzforchung*, 60(1), 29-31, 2006.
7. Ausloos, J., et al., "Designing-by-Debate: A Blueprint for Responsible Data-Driven Research & Innovation", in *Responsible Research and Innovation Actions in Science Education, Gender and Ethics*, Springer, 2018.
8. Aydin, E.S., Yucel, O., Sadikoglu, H., "Experimental study on hydrogen-rich syngas production via gasification of pine cone particles and wood pellets in a fixed bed downdraft gasifier", *International Journal of Hydrogen Energy*, 44(32), 17389-17396, 2019.
9. Azadeh, A., Arani, H.V., Dashti, H., "A stochastic programming approach towards optimization of biofuel supply chain", *Energy*, 76, 513-525, 2014.
10. Balabin, R.M., Safieva, R.Z., "Gasoline classification by source and type based on near infrared (NIR) spectroscopy data", *Fuel*, 87(7), 1096-1101, 2008.
11. Bassett, K., Liang, C., Marchessault, R., "The infrared spectrum of crystalline polysaccharides. IX. The near infrared spectrum of cellulose", *Journal of Polymer Science Part A: General Papers*, 1(5), 1687-1692, 1963.
12. Bellazzi, R., Zupan, B., "Predictive data mining in clinical medicine: current issues and guidelines", *International journal of medical informatics*, 77(2), 81-97, 2008.
13. Berndes, G., Hoogwijk, M., Van den Broek, R., "The contribution of biomass in the future global energy supply: a review of 17 studies", *Biomass and bioenergy*, 25(1), 1-28, 2003.
14. Breiman, L., "Bagging predictors", *Machine learning*, 24(2), 123-140, 1996.
15. Castillo, R., et al., "Supervised pattern recognition techniques for classification of Eucalyptus species from

- leaves NIR spectra", *Journal of the Chilean Chemical Society*, 53(4), 1709-1713, 2008.
16. Cortes, C., Vapnik, V., "Support-vector networks", *Machine learning*, 20(3), 273-297, 1995.
 17. Freund, Y., Schapire, R.E. Schapire R: Experiments with a new boosting algorithm. in *In: Thirteenth International Conference on ML*. Citeseer, 1996.
 18. Fujimoto, T., et al., "Application of near infrared spectroscopy for estimating wood mechanical properties of small clear and full length lumber specimens", *Journal of Near Infrared Spectroscopy*, 16(6), 529-537, 2007.
 19. Fujimoto, T., Yamamoto, H., Tsuchikawa, S., "Estimation of wood stiffness and strength properties of hybrid larch by near-infrared spectroscopy", *Applied spectroscopy*, 61(8), 882-888, 2007.
 20. Guenther, N., Schonlau, M., "Support Vector Machines", *The Stata Journal: Promoting communications on statistics and Stata*, 16(4), 917-937, 2016.
 21. Khan, F.M., Zubek, V.B. "Support vector regression for censored data (SVRc): a novel tool for survival analysis". in *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008.
 22. Kuo, P.-C., Wu, W., Chen, W.-H., "Gasification performances of raw and torrefied biomass in a downdraft fixed bed gasifier using thermodynamic analysis", *Fuel*, 117, 1231-1241, 2014.
 23. Lehmann, J., Joseph, S., "Biochar for environmental management: an introduction", in *Biochar for environmental management*, Routledge. p. 33-46, 2015.
 24. Lestander, T.A., et al., "Characterization of fast pyrolysis bio-oil properties by near-infrared spectroscopic data", *Journal of Analytical and Applied Pyrolysis*, 133, 9-15, 2018.
 25. Luypaert, J., Massart, D., Vander Heyden, Y., "Near-infrared spectroscopy applications in pharmaceutical analysis", *Talanta*, 72(3), 865-883, 2007.
 26. Ma, T., et al., "Rapid identification of wood species by near-infrared spatially resolved spectroscopy (NIR-SRS) based on hyperspectral imaging (HSI)", *Holzforschung*, 73(4), 323-330, 2019.
 27. Melkior, T., et al., "NMR analysis of the transformation of wood constituents by torrefaction", *Fuel*, 92(1), 271-280, 2012.
 28. Mitsui, K., Inagaki, T., Tsuchikawa, S., "Monitoring of hydroxyl groups in wood during heat treatment using NIR spectroscopy", *Biomacromolecules*, 9(1), 286-288, 2007.
 29. Mohammadi, K., et al., "Support vector regression based prediction of global solar radiation on a horizontal surface", *Energy Conversion and Management*, 91, 433-441, 2015.
 30. Mutlu, A.Y., Yucel, O., "An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification", *Energy*, 165, 895-901, 2018.
 31. Nefeslioglu, H., et al., "Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey", *Mathematical Problems in Engineering*, 2010.
 32. Porep, J.U., Kammerer, D.R., Carle, R., "On-line application of near infrared (NIR) spectroscopy in food production", *Trends in Food Science & Technology*, 46(2), 211-230, 2015.
 33. Rokach, L., "Ensemble-based classifiers", *Artificial Intelligence Review*, 33(1-2), 1-39, 2010.
 34. Rousset, P., et al., "Characterisation of the torrefaction of beech wood using NIRS: Combined effects of temperature and duration", *Biomass and bioenergy*, 35(3), 1219-1226, 2011.
 35. Safarian, S., Unnþórsson, R., Richter, C., "A review of biomass gasification modelling", *Renewable and Sustainable Energy Reviews*, 110, 378-391, 2019.
 36. Sandak, A., Sandak, J., Negri, M., "Relationship between near-infrared (NIR) spectra and the geographical provenance of timber", *Wood science and technology*, 45(1), 35-48, 2011.
 37. Schimleck, L., et al., "Comparison of methods for estimating mechanical properties of wood by NIR spectroscopy", *Journal of Spectroscopy*, 2018.
 38. Schwanninger, M., et al., "Application of Fourier transform near infrared spectroscopy (FT-NIR) to thermally modified wood", *Holz als Roh-und Werkstoff*, 62(6), 483-485, 2004.
 39. Schwanninger, M., Rodrigues, J.C., Fackler, K., "A review of band assignments in near infrared spectra of wood and wood components", *Journal of Near Infrared Spectroscopy*, 19(5), 287-308, 2011.
 40. So, C.-L., et al., "Near Infrared Spectroscopy in the Forest Products Industry", *Forest Products Journal*, *Forest Products Journal*, 54(3), 6-16, 2004.
 41. Solomon, S., et al., "Irreversible climate change due to carbon dioxide emissions", *Proceedings of the National Academy of Sciences*, 106(6), 1704-1709, 2009.
 42. Tang, Z., Maclennan, J., "Data mining with SQL Server 2005". *John Wiley & Sons*, 2005.
 43. Tsuchikawa, S., "A review of recent near infrared research for wood and paper", *Applied Spectroscopy Reviews*, 42(1), 43-71, 2007.
 44. Tsuchikawa, S., Yonenobu, H., Siesler, H., "Near-infrared spectroscopic observation of the ageing process in archaeological wood using a deuterium exchange method", *Analyst*, 130(3), 379-384, 2005.
 45. van der Ploeg, T., Austin, P.C., Steyerberg, E.W., "Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints", *BMC medical research methodology*, 14(1), 137, 2014.
 46. Vapnik, V., "The nature of statistical learning theory", *Springer science & business media*, 2013.
 47. Via, B.K., Adhikari, S., Taylor, S., "Modeling for proximate analysis and heating value of torrefied biomass with vibration spectroscopy", *Bioresource technology*, 133, 1-8, 2013.
 48. Workman Jr, J., Weyer, L., "Practical guide and spectral atlas for interpretive near-infrared spectroscopy", *CRC press*, 2012.

49. Xu, M., et al., "Decision tree regression for soft classification of remote sensing data", *Remote Sensing of Environment*, 97(3), 322-336, 2005.
50. Yang, H., Sheng, K., "Characterization of biochar properties affected by different pyrolysis temperatures using visible-near-infrared spectroscopy", *ISRN Spectroscopy*, 2012.
51. Yeh, T.-F., Chang, H.-m., Kadla, J.F., "Rapid prediction of solid wood lignin content using transmittance near-infrared spectroscopy", *Journal of Agricultural and food chemistry*, 52(6), 1435-1439, 2004.
52. Yucel, O., Aydin, E.S., Sadikoglu, H., "Comparison of the different artificial neural networks in prediction of biomass gasification products", *International Journal of Energy Research*, 43(11), 5992-6003, 2019.