Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Multiclass moisture classification in woodchips using IIoT Wi-Fi and machine learning techniques

Kerul Suthar, Q. Peter He*

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA

ARTICLE INFO

Article history:
Received 4 February 2021
Revised 12 June 2021
Accepted 5 July 2021
Available online 20 July 2021

Keywords:
Woodchip moisture content
Machine learning
Internet-of-things
Channel state information
Advanced manufacturing
Big data analytics

ABSTRACT

For the pulping process in a pulp & paper plant that uses woodchips as raw material, the moisture content (MC) of the woodchips is a major process disturbance that affects product quality and consumption of energy, water and chemicals. Existing woodchip MC sensing technologies have not been widely adopted by the industry due to unreliable performance and/or high maintenance requirements that can hardly be met in a manufacturing environment. To address these limitations, we propose a non-destructive, economic, and robust woodchip MC sensing approach utilizing channel state information (CSI) from industrial Internet-of-Things (IIoT) based Wi-Fi. While these IIoT devices are small, low-cost, and rugged to stand for harsh environment, they do have their limitations such as the raw CSI data are often very noisy and sensitive to woodchip packing. To address this, statistics pattern analysis (SPA) is utilized to extract physically and/or statistically meaningful features from the raw CSI data, which are sensitive to woodchip MC but not to packing. The SPA features are then used for developing multiclass classification models using various linear and nonlinear machine learning techniques to provide potential solutions to woodchip MC estimation for the pulp and paper industry. This work also demonstrates that classification accuracy alone is not a good performance metric for industrial applications, and the practical implications of misclassification must also be considered.

1. Introduction

The US is one of the largest producers of pulp products as well as one of the largest producers of paper and paperboard products. The US pulp and paper industry ranks the third in terms of energy consumption among US industries and spends over $7 billion annually on purchased fuels and electricity (Kramer et al., 2011). The pulping process, which converts woodchips into pulp by displacing lignin from cellulose fibers, is one of the most energy intensive processes and has been identified by the ENERGY STAR® and the Department of Energy (DOE) reports as a major opportunity to improve energy productivity and efficiency of the industry (Brueske et al., 2015; Kramer et al., 2011; Martin et al., 2000). Currently, vast majority of the US pulp is produced by chemical pulping processes and most of them utilize continuous Kamyr digesters. A Kamyr digester is a complex vertical plug flow reactor where the woodchips react with an aqueous solution of sodium hydroxide and sodium sulfide, also known as white liquor, at elevated temperatures to remove lignin. For Kamyr digesters, the incoming woodchip moisture content (MC) is a major source of disturbance that affects the cooking performance, as it dilutes the white liquor concentration therefore reducing the delignification reaction rate. In this work, wet basis MC is used, which is defined as the following:

$$MC = \frac{m_W}{m_T} \times 100\% = \frac{m_W}{m_W + m_D} \times 100\% \tag{1}$$

where m_W, m_D, and m_T represent the mass of water, dry wood, and total mass, respectively. Currently, the woodchip MC is not measured in real-time due to the lack of affordable, reliable, and easy-to-maintain sensors (Rahman et al., 2020). Instead, woodchip MC is commonly measured only four times per year corresponding to the four seasons and used to determine the operation parameters such as chemical usage and cooking temperature. Because this significant process disturbance is unmeasured, the performance of existing control solutions is often unsatisfactory and process engineers often overcook the woodchips to ensure pulp quality, which results in significant loss of pulp yield, overuse of heat/energy and chemicals. Chemical overuse also adds burdens to the downstream processes, such as washing and evaporation, and results in increased energy and chemical usages for downstream processes as well (Z. Jiang at the Alabama Center for Paper and Bioresource Engineering, personal communication, 2020). It is worth noting that there have been significant efforts and ad-

* Corresponding author.
E-mail address: qhe@auburn.edu (Q.P. He).

https://doi.org/10.1016/j.compchemeng.2021.107445
0098-1354/© 2021 Elsevier Ltd. All rights reserved.

vancements in the modeling and control of chemical pulping over the past decade (Rahman et al., 2020). In particular, progress has been made on multiscale modeling of Kraft pulping processes to capture the evolution of fiber morphology such as fiber length, porosity, and cell wall thickness (CWT) of cooked pulp (Choi and Kwon, 2019a; Choi and Kwon, 2019b). A recent study integrates macroscopic and microscopic models of the Kraft pulping process to develop an inferential model predictive control (MPC) for better handling of pulp grade transitions (Choi et al., 2021). These efforts have not explicitly considered the woodchip MC variability in a production environment, and this information, if made available, can be directly incorporated into these models for improved model accuracy in practical applications.

The oven-drying method is a direct and precise method based on the weight loss after a drying process, with standard defined by American Society for Testing and Materials (ASTM) (ASTM, 2016; Reeb and Milota, 1999). However, it is an offline test that takes 24 hours, and is mainly used for validating other indirect methods. A variety of indirect sensing methods have been examined for measuring woodchip MC online, including technologies that are based on microwave (Daassi-Gnaba et al., 2018), radio-frequency (RF) (Daassi-Gnaba et al., 2017), capacitance (Fridh et al., 2018; Pan et al., 2017, 2016), Resonant half-wave antenna (Merlan et al., 2019), near-infrared (NIR) (Amaral et al., 2020; Liang et al., 2019) and X-ray (Couceiro et al., 2019; Hultnäs and Fernandez-Cano, 2012). However, these methods have not been widely adopted by the industry due to poor performance and/or high maintenance requirements that can hardly be met in a manufacturing environment.

To address the robustness and performance limitations of the existing methods, we propose a non-destructive, economic, and robust approach based on 5 GHz IIoT short-range Wi-Fi and use channel state information (CSI) to predict MC in woodchips. CSI data have been used for moisture and mildew detections in wheat (Hu et al., 2019a; Yang et al., 2018a,b). However, woodchip MC classification is a much more challenging task due to the much bigger size and significantly more heterogeneous in both size and shape of the woodchips than those of wheat. Because of that, the woodchip packing or arrangement in the container is expected to have a significant impact on the CSI data, *i.e.*, woodchip packing is a strong confounding factor to MC level. There are generally three ways to address confounding variables: elimination, measuring, and randomization. Since woodchip packing cannot be eliminated nor measured, randomization is the approach taken in this work to address it. In addition, our recent studies have shown that IIoT sensors have their own shortcomings, including significant noise, missing values, and/or irregular sampling intervals, which result in messy big data and lead to low performing models when directly fed to machine learning algorithms (Shah et al., 2019a, 2017). Because of these challenges, the normalized or principal component analysis (PCA) transformed raw CSI data, which were used for wheat MC classification, are no longer sufficient for woodchip MC classification. To address it, the statistics pattern analysis (SPA) framework that we developed previously (He and Wang, 2018a,b; Suthar et al., 2019; Wang and He, 2010) is used to extract robust and predictive features from the raw noisy CSI data. These features are shown to be sensitive to the MC levels but insensitive to the packing of the woodchips. It is worth noting that SPA features are physically and/or statistically meaningful while other algorithmically generated features (*e.g.*, square, square root, exponential, etc.) or kernel-type features are often unintuitive. SPA also eliminates the data preprocessing steps (*e.g.*, outlier detection and handling, environmental noise removal) that were required in previous studies (Hu et al., 2019b; Yang et al., 2018a,b). These two strategies utilized for addressing a confounding variable and for extracting predictive and meaningful features from raw CSI data are two of the main contributions of this work. Another contribution of this work is the systematic study of different state-of-the-art linear and nonlinear classification techniques, as well as individual vs. ensemble classification, for woodchip MC classification using CSI data. Finally, classification accuracy has been commonly used in previous studies for evaluating classifier performance. We show that the classification accuracy alone is not a good performance metric, and the practical implications (*e.g.*, cost) of misclassification must also be considered.

The remainder of the paper is organized as follows. A brief background on CSI and feasibility study for using CSI in woodchip MC detection are presented in Section 2. Section 3 outlines the experimental setup and data collection procedure. In Section 4, we discuss the challenge of classification using raw data and the need of feature engineering, followed by the proposed approach based on statistics pattern analysis (SPA) for feature extraction. The classification approaches studies in this work are introduced in Section 5, along with the hyperparameter optimization approach used in this work. In Section 6, the results from different classification techniques are discussed in terms of both classification accuracy and robustness. The practical implications of these results are also discussed. Finally, conclusion and future work are discussed in Section 7.

2. Channel state information and feasibility for woodchip MC classification

2.1. Channel state information (CSI)

Using Wi-Fi cards such as Intel Wi-Fi link 5300 network interface card (IWL5300 NIC), it is convenient to collect CSI measurements that record the channel variation during propagation of wireless signals. In this work, CSI is extracted by modifying the open source device drivers for IWL5300 based on CSITool (Halperin et al., 2011). Similar tools are available based on Atheros chipsets as well (Xie et al., 2018). CSI amplitude and phase data are collected in this work using IWL5300 NIC by configuring the transmitter and receiver in injection and monitor modes, respectively. We use Lenovo ThinkPad systems equipped with Linux based OS 14.02 and kernel version 4.2 due to the version-specific CSI tool. Both systems are equipped with IWL5300 NIC with a modified driver and firmware for data collection. Orthogonal frequency-division multiplexing (OFDM) is often utilized to deal with impairments in wireless propagation such as frequency selective fading. In OFDM signal modulation, a single data stream is split into multiple orthogonal subcarriers at different frequencies to avoid interference and crosstalk. The IWL5300 NIC used in this work implements an OFDM system with 56 subcarriers, out of which 30 subcarriers can be read using the CSItool, which is built on IWL5300 NIC using a custom modified firmware and open source Linux wireless drivers (Halperin et al., 2011). Each channel matrix entry is a complex number, with signed 8-bit resolution each for the real and imaginary parts. It specifies the gain and phase of the signal path between a single transmit-receive antenna pair. For example, the channel response of the i^{th} subcarrier can be represented as:

$$CSI_i = |CSI_i| \exp\{\angle CSI_i\} \tag{2}$$

where $|CSI_i|$ is the amplitude response of the i^{th} subcarrier and $\angle CSI_i$ is the phase response.

CSI can reflect indoor channel characteristics such as multipath effect, shadowing, fading, and delay (Ahamed and Vijay, 2017). Our hypothesis is that the water content in the woodchips has a detectable impact on the strength and/or the phase of the signals that are received on the receiver side. In other words, woodchips at different MC levels would lead to different characteristics of CSI signal in terms of amplitude and/or phase responses. There-

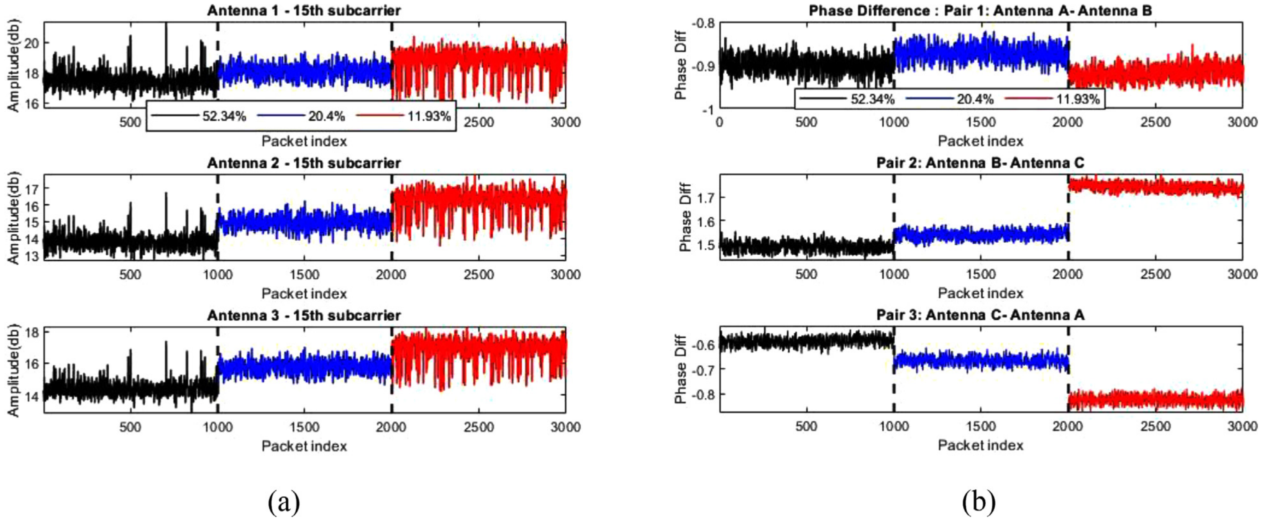


Fig. 1. CSI signals collected on the three receiving antennas at three different MC levels: (a) amplitude; (b) phase difference. Only signals from subcarrier 15 are shown.

fore, machine learning algorithms can be utilized to correlate these characteristics to woodchip MC levels.

In this work, two laptops equipped with IWL5300 NIC and modified drivers with specific Linux kernels are used to collect CSI data. One is set in injection mode while the other is set in monitor mode to collect 5 GHz CSI amplitude and phase data. One antenna is used on the transmitter side, while three antennas are used on the receiver side to take the advantage of the multiple-input multiple-output (MIMO) systems for improving diversity of signals (Ahamed and Vijay, 2017; Halperin et al., 2012). This diversity is exploited in this work to improve the multiclass classification performance. Also, it is desirable to focus the RF energy in one direction as the woodchips are placed in an airtight box between the transmitter and receiver. Therefore, unidirectional antennas are selected over omnidirectional antennas. As the gain of the directional antennas increase, the coverage distance also increases in that direction. Also, directional antennas for point-to-point connection reduce interferences from other sources. In this work, panel antennas ALFA (ALFA Network, Taiwan) with 66° horizontal beam-width and 16° vertical beam-width are used.

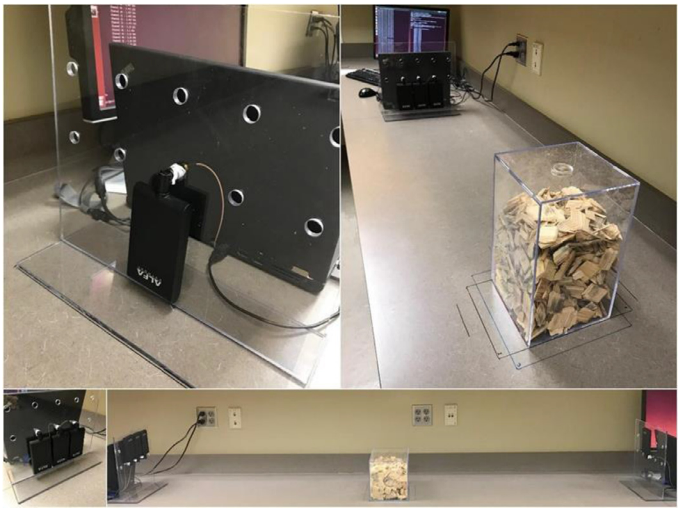


Fig. 2. Experimental setup for CSI data collection.

2.2. Feasibility test

To test the technical feasibility of CSI to classify woodchips based on MC levels, we collect CSI for woodchips at three distinctively different MC levels (*i.e.*, 52.34%, 20.40% and 11.93%). Fig. 1 shows the CSI amplitude and phase difference for the 15th subcarrier. As shown in Fig. 1, there are distinctive differences in amplitude and phase difference of different MC levels from all three antennas. This preliminary feasibility test indicates that it is possible to develop a data-driven model to correlate CSI data with woodchip MC level. Note that the confounding factor of woodchip packing is not considered here.

3. Experimental setup and data collection

3.1. Experimental setup

With the results from the feasibility test in Section II, we design an experimental setup with antenna positions fixed on an acrylic sheet. The experimental setup is shown in Fig. 2, where two Lenovo T400s systems equipped with IWL5300 NIC are set 3 m apart. The woodchips at different MC levels are placed at the center (*i.e.,* 1.5 m from transmitting and receiving antennas) in an acrylic container with an air-tight lid to avoid any changes in MC while the data are being collected.

3.2. Data collection

In previous studies a maximum of 5 MC levels have been studied with minimum difference of 0.7% in MC (Yang et al., 2018a). However, this is not nearly sufficient for woodchip MC levels because woodchips are usually stored outdoors, which introduces significant MC variations due to daily weather conditions, and seasonal temperature and humidity changes. In this work, data are collected for 20 different MC classes or levels ranging from 53.39 % to 11.81% on the wet basis (see Eq. (1)). Total mass (m_T) is measured during each experiment and oven drying method (ASTM, 2016; Reeb and Milota, 1999) was performed after all experiments were conducted to determine the oven dry weight (m_D). m_T and m_D are then used to determine the mass of water (m_W) and MC according to Eq. (1). The 20 different MC levels are plotted in Fig. 3. There are two gaps in the tested MC levels, one around 45% and the other around 25%. This is due to the overnight exposure of the woodchips to air in the lab, which should be avoided if a model to be developed for accurate estimation of any MC level in the entire range. Nevertheless, this does not affect our methodol-

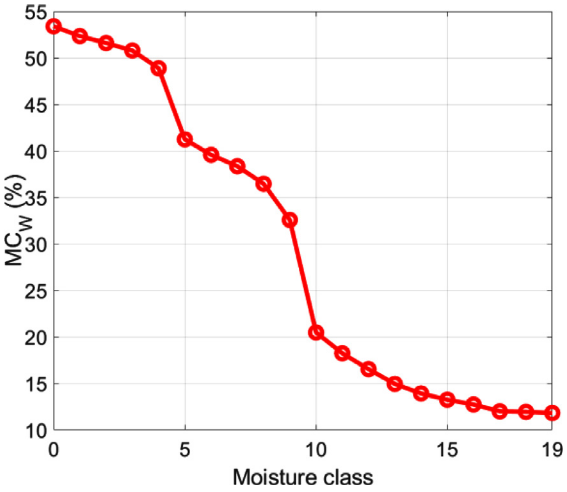


Fig. 3. 20 different MC classes/levels for experimental data collected.

ogy development, nor the conclusions drawn based on the results obtained. This is because there are three regions where MC levels are reasonably separated as shown in Fig. 3. In addition, MC levels are narrowly separated at the high MC region and even more so at the low MC region. The minimum difference between MC levels is 0.05%. If a model can correctly classify MC levels with such narrow difference, we expect it to work if more MC levels were included in the middle range with wider difference such as 1%, which is sufficient for pulping process optimization and control.

As discussed previously, the woodchip packing or arrangement in the container is expected to have significant impact on CSI data. Based on the principle of randomization for addressing this confounding factor, the woodchips within the air-tight box are shuffled 10 times for each MC level. In other words, for each MC level, 10 datasets (i.e., samples) are collected corresponding to 10 shuffles. Therefore, the experiments generate totally 200 samples for all 20 MC levels. For each sample, 10,000 packets were sent from the transmitter (setup in injection mode) to the three receiver antennas (setup in monitoring mode). Data are collected only for the line of sight (LOS) scenario, i.e., the woodchip container is placed in the middle of the center line between the transmitter and the receivers.

4. Feature engineering and selection

For wheat MC classification, normalized raw data were used in long short-term memory (LSTM) recurrent neural network (RNN) (Yang et al., 2018a) and RFB-NN (Hu et al., 2019a), while principal component scores from normalized raw data were used in support vector machines (SVM) (Yang et al., 2018b). In the next section, we show that raw data are poor features for woodchip MC classification due to the challenges discussed previously in Sec. 1. In addition, the Wi-Fi packets are independent from each other (*i.e.*, serially uncorrelated) as evidenced by the close-to-zero autocorrelation coefficients beyond lag 0. Therefore, there is no reason to use an RNN such as LSTM to account for the serial dependency or dynamics of packets.

4.1. The challenges of using raw CSI data as features

As discussed in the previous section, for each MC level we shuffle the woodchips 10 times and collect CSI data for each shuffle to address the confounding factor of woodchip arrangement or packing. Fig. 4 shows the raw CSI data of amplitude and phase difference for woodchips at five distinctively different MC levels with 10 shuffles at each MC level. The five MC levels are: 53.29%, 41.24 %, 32.57 %, 20.47 % and 11.81 %, in that order where they are plotted in Fig. 4. For the sake of better visualization and easier interpretation, only 100 packets from the 10th subcarrier for each shuffle are plotted.

From Fig. 4, the impact of shuffling can be seen in both amplitude and phase difference, although it is more obvious in the phase difference. The observation confirms our earlier suspicion that packing or woodchip arrangement is a significant confounding factor to MC level. In addition to packing, another challenge is the significant noises presented in both amplitude and phase difference. Finally, Fig. 4 shows no clear trend or pattern in amplitude or phase difference that correlates with MC levels. All these factors present significant challenges to model MC level with raw CSI data. As an illustrative example, we use linear discriminant analysis (LDA) to perform classification using the raw CSI data, with either amplitude, or phase difference, or both. For training, 9 samples are randomly selected from 10 shuffled samples at each of the 20 MC levels, which results in 180 training samples. The remaining one shuffled sample from each MC level is used for testing after the classification model is trained. This process is repeated 100 times, resulting in 100 Monte Carlo runs and the classification results are shown in Fig. 5. For performance evaluation, the classification accuracy of class i is defined as

$$Accuracy_i = \frac{p_i}{n_i} \tag{3}$$

The overall accuracy of all classes is defined as

$$Accuracy = \frac{\sum_{i=1}^{C} p_i}{\sum_{i=1}^{C} n_i} = \frac{\sum_{i=1}^{C} p_i}{N} \tag{4}$$

where C denotes total number of classes, n_i true/known number of samples in class i, $N = \sum_{i=1}^{C} n_i$ total number of samples, and p_i number of correctly predicted samples in class ... Fig. 5 (a) compares the overall classification accuracy of all classes when different components of the CSI data were used. The comparison indicates that LDA classifier using both amplitude and phase difference performs the best with 86.15% classification accuracy, followed by LDA classifier using phase difference with 83.85% classification accuracy, while the LDA classifier using amplitude alone results in the lowest classification accuracy of 76.10%. Fig. Fig. 5 (b) plots the confusion matrix for the best LDA classifier of using CSI amplitude and phase difference, which allows us to dig deeper into the classification results. As can be seen from Fig. 5 (b), classification accuracy of individual classes ranges from 15% to 100%. It can also be seen that classification accuracy alone is not a good performance indicator. For example, classification accuracy alone would not be able to distinguish the following two scenarios: (1) the actual scenario of misclassifying ten 53.38% MC level samples (class 0 in Fig. 5 (b)) to 16.52 % MC level (class 12); (2) a hypothetical scenario of misclassifying ten 53.38% MC level (class 0) to 51.59 % MC level (class 1). Both scenarios have a classification accuracy of 90%, but with drastically different implications in this application. For example, if MC level is used to control the chemical usage, the former would let to significantly worse outcome than the latter. With this point in mind, we see from Fig. 5 (b) that the classification results using raw CSI data are poor as there are samples misclassified far off their actual classes. In this work, when the predicted class of a sample is off its true class by more than one level, we term it “far-off misclassification”, to distinguish it from the scenario of “nearest-neighbor misclassification”, where the predicted class is off true class by one level (either above or below). Based on this definition, there are totally 478 misclassified samples, of which 30 are far-off misclassifications (highlighted by red circles in Fig. 5 (b)).

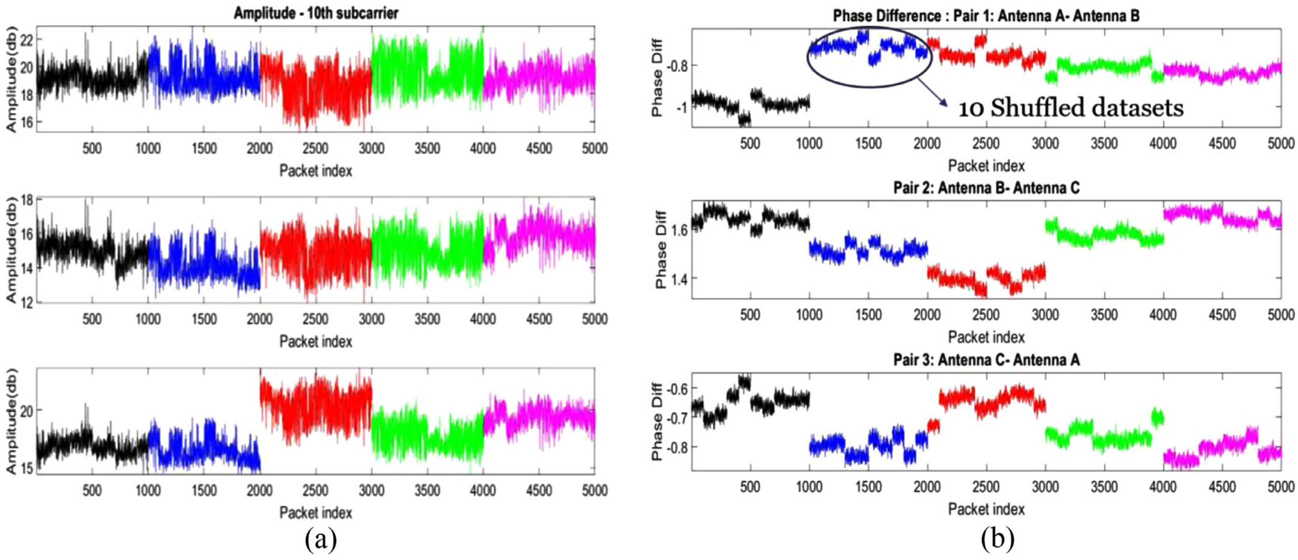
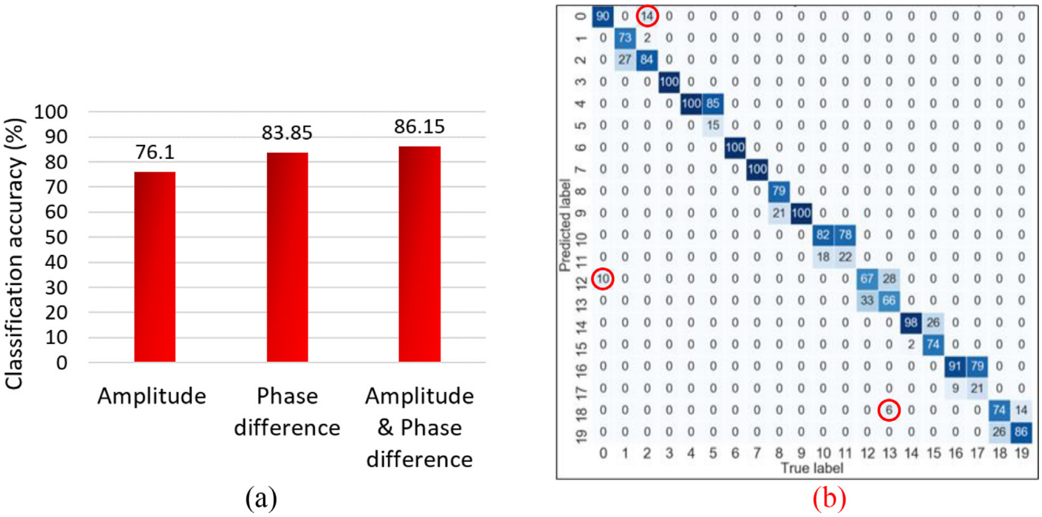


Fig. 4. Raw CSI data for 5 different MC levels showing 10 shuffles for each MC level: (a) amplitude; (b) phase difference.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	90	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	73	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	27	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	100	85	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	21	100	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	82	78	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	18	22	0	0	0	0	0	0	0	0
12	10	0	0	0	0	0	0	0	0	0	0	0	67	28	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	33	66	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	26	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	74	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	91	79	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	21	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	74	14
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	86

Fig. 5. (a) Overall classification accuracy using different raw CSI data with LDA classifier based on 100 Monte Carlo runs. (b) Classification confusion matrix of 100 Monte Carlo runs when both amplitude and phase difference are used. Since there are 100 samples in each class (true labels), the numbers on diagonal represent the percentage of classification accuracy of individual classes, which range from 15% to 100%. The far-off misclassifications (*i.e.*, the predicted class differs from the true class by more than one MC level) are highlited by red circles.

4.2. Feature engineering with statistics pattern analysis (SPA)

To address the shortcoming of raw CSI features that lead to not only low classification accuracy but also far-off misclassifications, in this work, statistics pattern analysis (SPA) is utilized to generate more robust and predictive features. SPA was proposed to supplement the traditional multivariate modeling approaches that directly utilize process variables (e.g., temperature, pressure, etc.) for monitoring, control and inference purposes. In SPA, the statistics of the process variables, instead of process variables themselves, are used for modeling. The statistics capture the characteristics of each individual variable (e.g., mean and variance), the interactions among different variables (e.g., covariance), the dynamics (e.g., auto-, cross-correlations), as well as process nonlinearity and process data non-Gaussianity (e.g., skewness, kurtosis, and other higher order statistics or HOS). SPA is based on hypothesis that these statistics are sufficient and even better in capturing process characteristics (e.g., static properties and dynamic behaviors) than original process variables. This hypothesis has been supported in various applications, including fault detection (He et al., 2019; He and Wang, 2018a, 2011; Wang and He, 2010), fault diagnosis (He et al., 2012; He and Wang, 2018a), and virtual metrology or soft sensor (Shah et al., 2020, 2019b; Suthar et al., 2019, 2018). Due to the fact that statistics are computed using a set of observatons, they are less affected by noises. In addition, there are robust statistics that are insensitive to outliers. Finally, due to central limit theorem (CLT), these statistics are asymptotically normally distributed. For these reasons, SPA is selected in this work to extract robust and predictive features from raw CSI data. It is worth noting that SPA does not require preprocessing of the CSI data (i.e., outlier detection and handling, noise removal/reduction) that has been required in previous studies (Hu et al., 2019b; Yang et al., 2018b,a). The schematic diagram of SPA based classification is shown in Fig. 6. In the first step, various statistics are extracted from the CSI amplitude and phase data.

$$\mathbb{P} : \boldsymbol{X} \rightarrow \boldsymbol{F} \tag{5}$$

Table 1
Statistics considered as features in this work.

Statistics	Definition	Statistics per sample
Mean	$\mu(x) = \frac{1}{K}\sum_{i=1}^{K} x_i$, where x is a CSI amplitude or phase difference variable	150
Median	$Med(x) = \frac{1}{2}(\vec{x}_{(K+1)/2} + \vec{x}_{(K+1)/2})$ where $\vec{x}$ denotes sorted x in ascending order; $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions respectively	150
Maximum	$Max(x) = \vec{x}_K$	150
Minimum	$Min(x) = \vec{x}_1$	150
Interquartile range	$IQR(x) = Q_3(x) - Q_1(x)$, where $Q_3(x)$ and $Q_1(x)$ are the upper and lower quartiles of x	150
Standard Deviation	$s(x) = \sqrt{\frac{1}{K-1}\sum_{i=1}^{K}(x_i - \mu(x))^2}$	150
Mean absolute deviation	$D_{mean}(x) = \frac{1}{K}\sum_{i=1}^{K}\lvert x_i - \mu(x)\rvert$	150
Median absolute deviation	$D_{med}(x) = \frac{1}{K}\sum_{i=1}^{K}\lvert x_i - Med(x)\rvert$	150
Coefficient of variation	$C_V(x) = \frac{s(x)}{\mu(x)}$	150
Skewness	$\gamma(x) = \frac{\frac{1}{K}\sum_{i=1}^{K}(x_i-\mu(x))^3}{s(x)^3}$	150
Kurtosis	$\kappa(x) = \frac{\frac{1}{K}\sum_{i=1}^{K}(x_i-\mu(x))^4}{s(x)^4}$	150
Cross-correlation coefficient (lag 0)	$R_{xy} = \frac{\frac{1}{K}\sum_{i=1}^{K}[(x_i-\mu(x))(y_i-\mu(y))]}{s(x)s(y)}$, where x and y are two CSI amplitude variables of the same antenna, or phase difference variables of the same antenna pair	$\frac{1}{2}(30 \times 29) \times 3 + \frac{1}{2}(30 \times 29) \times 2 = 2175$
Mean difference of consecutive subcarriers	$MDSC_{xy} = \mu(y) - \mu(x)$, where x and y are CSI amplitude or phase difference variables of two consecutive subcarriers of a same antenna	$29 \times 3 + 29 \times 2 = 145$

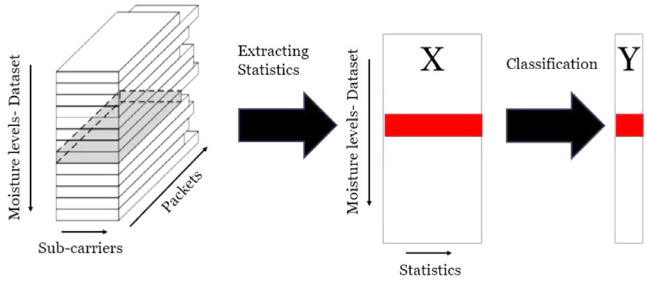


Fig. 6. Schematic of SPA-based feature extraction for classification.

where $\mathbb{P}$ denotes the operator that maps the 3D CSI data array $\boldsymbol{X} \in R^{N \times R \times K}$ containing N samples, R amplitudes and phase differences of all subcarriers from K packets into a feature matrix $\boldsymbol{F} \in R^{N \times S}$ containing N samples with each sample now characterized by S statistics, such as mean, standard deviation, skewness and kurtosis of amplitude of each subcarrier calculated over K packets. Note that K does not have to be the same across different samples, as long as it is sufficiently large to obtain reliable statistics. This is convenient if different number of packets were received for different samples. For between-variable statistics, between-subcarrier differences are considered, but between-packet statistics are not considered as packets are independent from each other. In Fig. 6, $\boldsymbol{Y} \in R^{N \times 1}$ denotes the MC levels for N samples. In the second step, a classification model is developed to capture the relationships between the sample features (*i.e.*, statistics) and the response (*i.e.*, MC levels). The SPA framework is a flexible method as different statistics can be added or removed based on how well they capture the relationships between the predictors and the response variables or classes.

Based on the SPA framework, we extracted 13 statistics (listed in Table 1) of 90 amplitude variables (i.e., 3 antennas, each with 30 subcarriers) and 60 phase difference variables (i.e., 2 independent antenna pairs, each with 30 subcarriers). All statistics are computed over 40,000 observations for each of the 200 samples (i.e., 10 samples/shuffles for each of the 20 MC levels). Autocorrelations are not considered because the packets are independent from each other as evidenced in Fig. 7 (a) where the sample autocorrelation coefficient of the CSI amplitude from one subcarrier of one antenna is shown, which resembles the pattern of a typical random signal. For cross-correlations, only cross-correlations between subcarriers of the same antenna with lag 0 is considered due to the absence of serial correlation between lags. Fig. 7 (b) shows the cross-correlation coefficient of CSI amplitude among subcarriers of a same antenna. It can be seen that CSI amplitude (and phase difference, not shown) from different subcarriers are highly correlated, especially the neighboring subcarriers. Because of this observation, we also considered mean difference between consecutive subcarriers. The idea is to capture the relationships between consecutive subcarriers in a more quantitatively way than cross-correlation coefficient between them. In this way, the overall shape of the CSI amplitude or phase difference across subcarriers can be captured.

Table 1 shows that there are 3,970 feature candidates for each sample, which is a rather large feature pool considering that we only have 200 samples. Therefore, a feature selection is desired before modeling to avoid over-fitting. There are many feature selection methods available. In this work we employ principal component analysis (PCA) for its simplicity and easy visualization, which is detailed in the next section.

4.3. Feature selection with principal component analysis (PCA)

The goal of feature selection is to find features that maximize between-class variance (i.e., distinct difference for samples of different MC levels) while minimizing within-class variance (i.e., high similarity for samples of the same MC level). For simplicity and robustness of features, we compare features by types listed in Table 1. This is conducted via unsupervised learning of PCA on each feature type and project them onto low-dimensional principal component subspace (PCS). Each feature was normalized to zero mean unit variance across all 200 samples prior to PCA. The results are illustrated in Fig. 8 where the 87 CSI amplitude mean difference of consecutive subcarriers (MDCSs) of 70 samples were

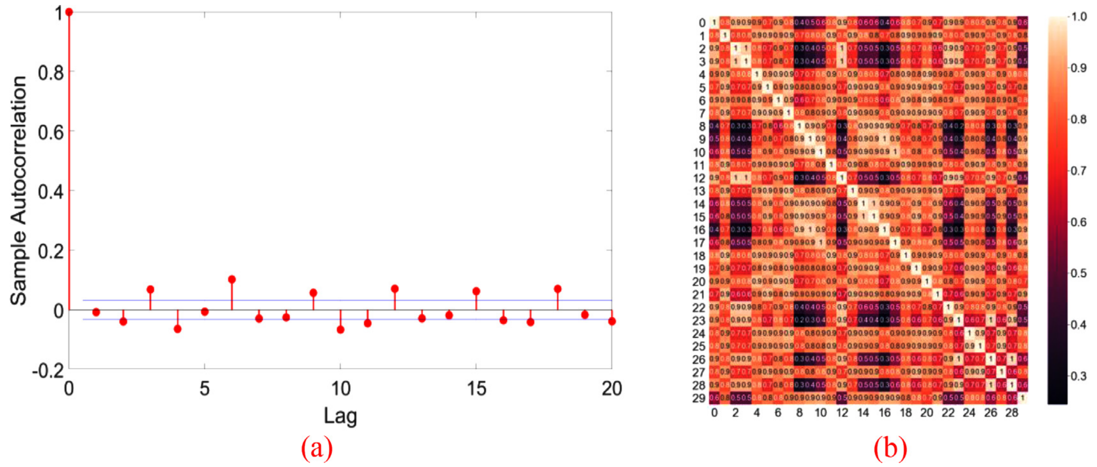


Fig. 7. (a) Auto-correlation coefficients of CSI amplitude of one antenna subcarrier over 40,000 packets show no significant serial correlation among packets; (b) Cross-correlation coefficients of CSI amplitude between subcarriers of one antenna show high correlations, especially between consecutive subcarriers.

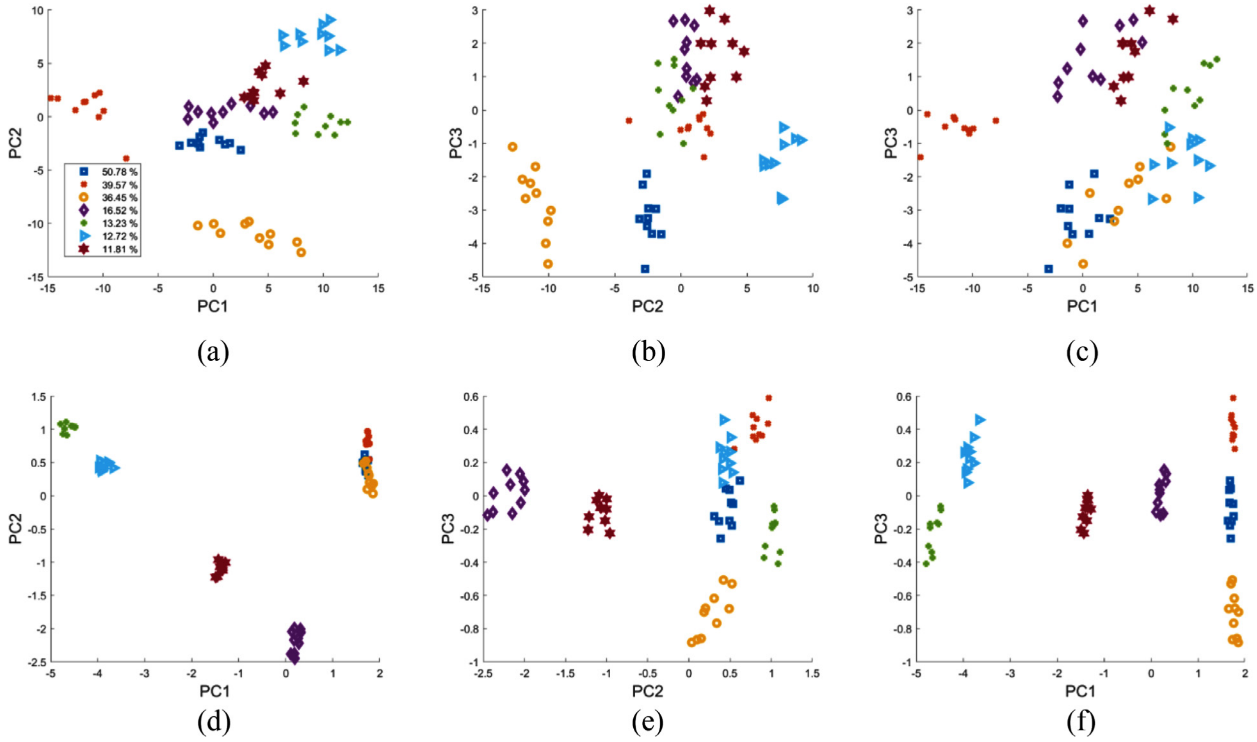


Fig. 8. (a)~(c): PCA score plots of CSI amplitude means of 70 samples at 7 different MC levels (i.e., 10 samples at each MC level); (d)~(f): PCA score plots of CSI amplitude mean difference of consecutive subcarriers (MDCS) of the same 70 samples. MDCSs show much better quality as features in both maximizing between-class variance and minimizing within-class variance.

projected onto the first three principal component directions to obtain the three "score" plots (Fig. 8 (d)~(f)). For comparison, the score plots of 150 CSI amplitude means of the same 70 samples were also generated (Fig. 8 (a)~(c)). As can be seen from Fig. 8, MDCSs show not only significant between-class differences (i.e., samples from different MC levels are far apart in one or multiple score plots), but also significant within-class similarities (i.e., samples from the same MC level but different shuffles form a compact cluster). In contrast, the mean of CSI amplitude is much more sensitive to woodchip packing, indicated by the wide scattering of samples from the same MC level but different shuffles. In addition, compared to CSI amplitude MDCS, the CSI amplitude mean is less sensitive to MC levels, indicated by the less separation of samples from different MC levels. As mean directly resembles raw data behavior, this is an indication of potentially poor performance for classification using raw data, which was verified in the previous section. Through this comparison of all feature types listed in Table 1, it was found that the MDCSs of CSI amplitude are the best feature candidates and therefore were selected as the final features for developing classification models. In this way, we reduce the feature space from 3,970 to 87. It is worth noting that further feature selection can be conducted to use MDCSs of selected subcarriers instead of all 30 subcarriers. It is also worth noting that classification performance is expected to improve if more a systematic feature selection is conducted, such as combining features from different types. These will be our future work to further improve the technology. However, in this work, we try to strike a bal-

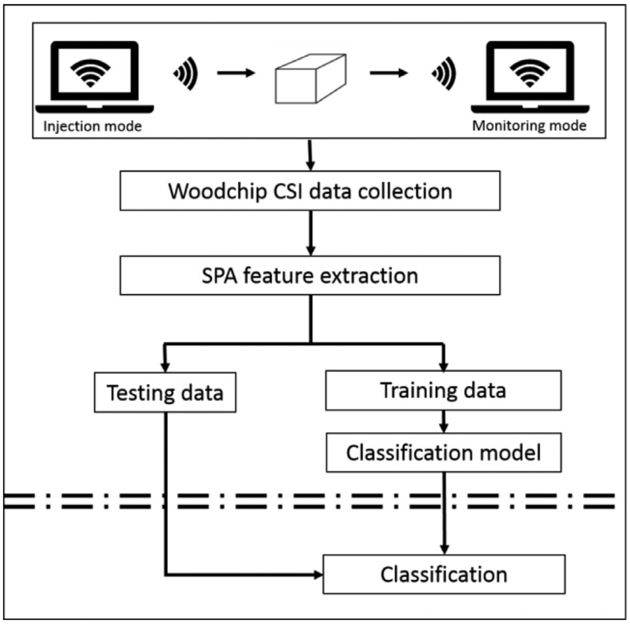


Fig. 9. Overall process flow diagram of woodchip MC level classification using CSI data.

ance that leans more towards simplicity and robustness than numerical performance.

5. Model building

Once the 87 CSI amplitude MDCSs are selected as the features, the next step is to develop classification models. In this work we compare various state-of-the-art machine learning classification techniques in classifying woodchip MC levels using these features. The procedure is outlined in Fig. 9. For each classification model, 9 samples are randomly selected from 10 shuffled samples at the same MC level for each of the 20 MC levels, which results in 180 training samples. The remaining one shuffled sample from each MC level is used for independent testing, which results in a total of 20 testing samples. Due to the limited number of samples, a Monte Carlo validation and testing (MCVT) procedure (Shah et al., 2019b) is followed to repeat the random sample selection, and model training and testing procedure 100 times. In addition to the mean and standard deviation of the overall classification accuracy (Eq. (4)) of 100 such MCVTs, the confusion matrix resulted from the same MCVTs is also used to evaluate the performance of different classification models.

The machine learning classification techniques studied in this work include linear discriminant analysis (LDA), support vector machine (SVM), artificial neural network (ANN), as well as ensemble modeling of bagging with LDA, and ensemble boosting method XGBoost. These methods are briefly reviewed in the following sections.

5.1. Linear discriminant analysis (LDA)

LDA is a robust supervised learning technique for multiclass classification. It is a generalization of Fisher's linear discriminant which find a linear combination of features to separate multiples classes in the dimensional space. Scikit-learn Python library (Pedregosa et al., 2011) is used to implement LDA in this work, which fits a Gaussian density to each class and estimates the class conditional distribution of data for each class k using Bayes' theorem:

$$P(y=k|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|y=k)P(y=k)}{P(\boldsymbol{x})} = \frac{P(\boldsymbol{x}|y=k)P(y=k)}{\sum_{l=1}^{C}\{P(\boldsymbol{x}|y=l)P(y=l)\}} \tag{6}$$

where $\boldsymbol{x} \in R^d$ is a sample feature vector of dimension d, y is the class label of that sample, C is the total number of classes. LDA makes predictions by estimating the probability of a new sample belonging to each class. Based on the class with the highest probability, the new sample is assigned to that class. More information on multiclass LDA can be found in (Friedman et al., 2001).

5.2. Support vector machine (SVM)

Support vector machine (SVM) is a supervised machine learning technique. In linear SVM classification of two classes, classification is performed by finding a hyperplane that maximizes the separation or margin between the two classes. If the two classes are not linearly separable, the input vectors can be nonlinearly mapped to a high-dimensional feature space through a kernel function that presumably makes the separation easier in the kernel space. In this application, it was found that linear SVM performs better than nonlinear kernel (e.g., radial basis function (RBF) and sigmoid kernels) based SVMs. This is consistent with the preliminary finding in the previous section where a subset of 7 classes were shown to be linearly separable (Fig. 8 (d)~(f)). More information on SVM can found in (Bishop, 2006; Burges, 1998; Cortes and Vapnik, 1995). In this work, multiclass classification is carried out using scikit-learn (Pedregosa et al., 2011) with the "one-versus-one" approach where 190 (i.e.,$(20 \times 19)/2$) classifiers are constructed.

5.3. Artificial neural network (ANN)

Artificial neural network (ANN), or simply neural network (NN), was developed with the idea of mimicking human brains, which now form the foundation of many deep learning techniques. A neural network consists of several layers, including an input layer that takes input data, one or more hidden layers depending on the complexity of the problem and the representations to be learned, and an output layer that outputs either discrete or continuous values depending on the type of problem, i.e., classification or regression. The constructed ANN represents interconnected input and out units or nodes (called neurons), in which each connection (called an edge) has an associated weight. The training of an ANN for classification is to adjust these weights to optimize the prediction of correct classes for the training data (e.g., through minimizing a cost function such as classification error). Once trained, the ANN takes a new set of similar data and make class predictions based on the trained model. Keras is used for ANN implementation in this work. Because of the likely linear separability of this particular application, one hidden layer is used in this work. Other hyperparameters, including number of neurons in the hidden layer, optimizer, activation function in the hidden layer, initialization, epochs and batch size, are optimized using random search followed by Bayesian optimization. More information on ANN can be found in (Aggarwal, 2018; Nielsen, 2016; Ripley, 1996; Theodoridis, 2015).

5.4. Bagging

Bagging is a bootstrap ensemble method that creates individual models for its ensemble by training each classifier on a random distribution of the training data. Each classifier's training set is generated by random sampling, with or without replacement from all the samples available for training. Individual predictions of each classifier are aggregated based on a voting scheme (hard voting or soft voting) to form a final prediction. A simple schematic

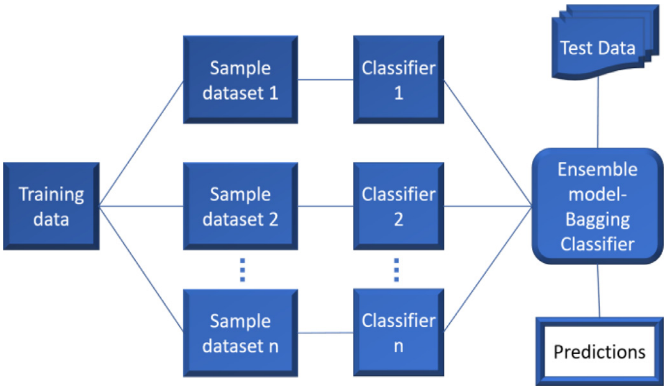


Fig. 10. Schematic of bagging-based classifier.

of a bagging classifier is shown in Fig. 10. Each base classifier can be trained in parallel with the sub samples generated with random sampling. Bagging is known to reduce overfitting or high variance by voting. Different base estimators can be used within bagging. In this work, LDA classifier is used due to the linear separability of the classes. Scikit-learn is used to implement bagging. The hyperparameters, including number of base classifiers, bootstrapping samples and/or features, and the sample/feature size, are optimized using random search followed by Bayesian optimization. More information on bagging can be found in (Breiman, 1999, 1996; Ho, 1998; Louppe and Geurts, 2012).

5.5. XGBoost

Another ensemble method that constructs multiple regression trees is boosting. In comparison to bagging, boosting approaches combine weak learners into a strong learner iteratively by optimizing a cost function along the negative gradient direction. XGBoost is one of the most successful boosting approaches under the gradient boosting framework. The XGBoost algorithm objective combines training loss and regularization terms for a trade-off on bias and variance. Python library xgboost is used for implementation. The hyperparameters, include number of base learners (i.e., regression trees), learning rate, updater, feature selector, and regularization parameters, are optimized using random search followed by Bayesian optimization. More information on XGBoost can be found in (Chen and Guestrin, 2016).

5.6. Hyperparameter optimization

Hyperparameter optimization is very important in training machine learning models as the model architecture directly affects the model performance. There are three major approaches for hyperparameter optimization, including grid search, random search (Bergstra and Bengio, 2012; Bergstra et al., 2011) and Bayesian optimization (Bergstra et al., 2013; Komer et al., 2014). Grid search can be quite effective when dealing with a small hyperparameter space. In general, however, random search and Bayesian optimization are more efficient than grid search. For complex models with large parameter spaces, such as XGBoost and ANN, the time required for grid search could be prohibitive. In these cases, random search or Bayesian optimization is preferred. Random search samples random parameter combinations based on certain statistical distributions. The idea is that, provided enough iteration, random search can find an optimum or close to optimum in lesser time than grid search, although random search does not guarantee a global optimum. Both grid search and random search find optimal hyperparameters in an isolated way without considering past evaluations. In contrast, Bayesian optimization considers past hyperparameter values that minimize the cost function by building a surrogate model based on past evaluation results. The surrogate model is presumably computationally cheaper to optimize than the original objective function, so the next input values are selected by applying criteria, such as expected improvement (EI), to the surrogate model. In this work, random search is utilized to explore the hyperparameter space. The final hyperparameters are determined by Bayesian optimization with Tree Parzen estimator using EI as the criterion. The Scikit-learn library is used for random search while hyperopt (Bergstra et al., 2013) is used for Bayesian optimization.

Table 2
Overall classification accuracy of LDA when features from single or all antennas are used.

Data used	Classification accuracy
Antenna 1	93.05 ± 5.17
Antenna 2	92.6 ± 5.97
Antenna 3	96.35 ± 3.40
All	97.55 ± 2.89

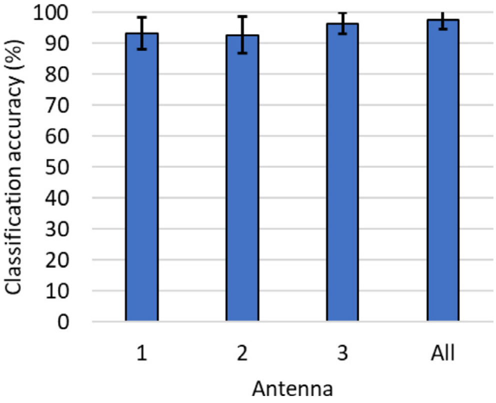


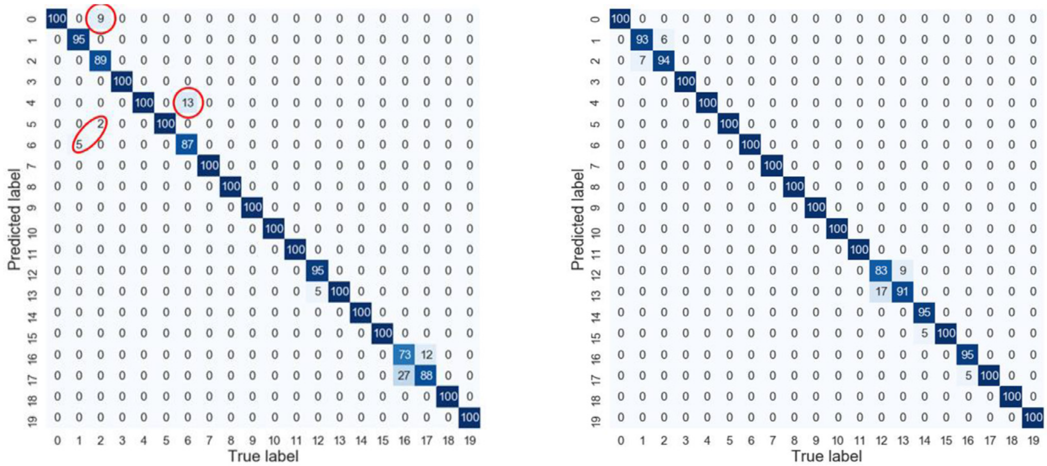
Fig. 11. Comparison of classification accuracy of LDA when features from different antennas are used.

6. Results and discussion

In this section, we discuss our findings of woodchip MC level classification using the 87 features extracted following the SPA framework. The classification results from the five different classification methods discussed in the previous section are compared. As discussed previously, due to the limited number of samples, 100 MCVT simulations are conducted. For every classification technique in each MCVT simulation, hyperparameters are optimized using stratified 10-fold cross validation. The trained model is used for evaluation on the set-aside testing set. The average and standard deviation of classification accuracy of 100 such runs (100 different test sets) are used to evaluate the performance of each classification method. In addition, the overall classification confusion matrix from 100 MCVTs is used to visualize and detect the far-off misclassifications where the predicted class is off its true class by more than one MC level.

We first investigate effect of antennas using LDA. The mean and standard deviation of overall classification accuracy are shown in Table 2 and Fig. 11. It can be seen that, when a single antenna (i.e., with 29 out of 87 features) is used, the antenna 3 provides the best information for classification. The best results, in both mean and standard deviation of classification accuracy, are obtained when all three antennas (i.e., with all 87 features) are used.

Another advantage of using all three antennas is observed when comparing the confusion matrix of different scenarios. Fig. 12 compares the confusion matrices of using only antenna 3 with that of using all three antennas. It can be seen that there are 29 far-off



(a)

(b)

Fig. 12. Comparison of classification confusion matrices when only features from antenna 3 are used (a) vs. when all features from all three antennas are used (b). The red circles/ellipse highlight the far-off misclassifications.

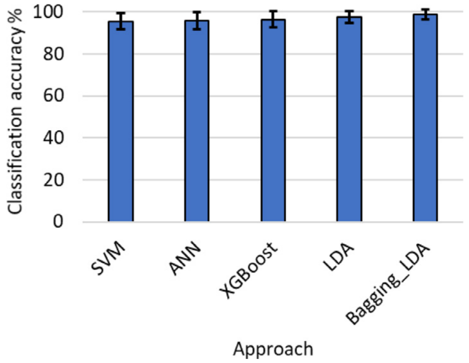


Fig. 13. Comparison of overall classification accuracy when different classification techniques are used.

Table 3
Overall classification accuracy when different classification techniques are used.

Method	Classification accuracy
SVM	95.50 ± 3.79
ANN	95.85 ± 4.15
XGBoost	96.40 ± 3.70
LDA	97.55 ± 2.89
Bagging (LDA)	98.75 ± 2.29

misclassifications when only antenna 3 is used. When all three antennas are used, there is no far-off misclassification occur. Therefore, for the remaining of this work, all 87 features from all three antennas are used.

Next, using all 87 features from all three antennas, we compare performance of different classification methods. The results are shown in Fig. 13 and Table 3, which indicate that all methods perform well and all achieve greater than 95% overall classification accuracy. SVM performs the worst among all methods in terms of mean classification accuracy. ANN performs slightly better than SVM in mean classification accuracy but with slightly higher standard deviation, indicating lower consistency when different training and testing samples are used. However, an analysis into the confusion matrices shows that SVM results in seven far-off misclassifications while all other methods result in zero far-off misclassification (Fig. 14). XGBoost performs slightly better than ANN and SVM, but not as good as LDA. This result is somewhat surprising as XGBoost has outperformed other techniques in many Kaggle competitions on real world datasets and a variety of applications. However, as shown earlier in Fig. 8, this application is more of a linearly separable case with the features selected, which explains the good performance from LDA. The results also demonstrate the robustness of LDA when dealing with linearly separable cases. Nevertheless, bagging can still improve a base classifier such as LDA in this work. As shown in Table 3, bagging of LDA provides the best performance with the highest average overall classification accuracy of 98.75% and the smallest standard deviation of 2.29% from 100 MCVT's. The confusion matrices of all methods, shown in Fig. 12 (b) for LDA and Fig. 14 for the other four methods, indicate that only SVM results in far-off misclassifications while all other methods only result in nearest-neighbor misclassification. The specific number of the two types of misclassifications are compared in Fig. 15, where the LDA on raw CSI amplitude data is used as the reference. Fig. 15 shows that feature engineering and selection plays a key role in this application, and all methods based on the 87 CSI amplitude MDCS features easily outperform LDA with raw CSI amplitude data as features.

We also compared the following two scenarios of hyperparameter tuning:

(A) A set of hyperparameters are optimized for each MCVT run using the selected training samples, and that set of hyperparameters are used for evaluation on the corresponding test set. Therefore, different MCVT runs could potentially have different hyperparameter values.

(B) The optimal hyperparameters from 100 MCVT's of Scenario A are stored and the mode of each hyperparameter (i.e., the value appears most frequently) is selected to construct a universal set of hyperparameters. The universal hyperparameter set is used for model training and testing of the same 100 sets of training and testing samples as in Scenario A.

One potential issue with Scenario B is that a test sample in one MCVT is potentially used as a training sample in other MCVT's. When the hyperparameters from all MCVT's are pooled together to determine the mode, essentially all samples have been used as training samples for hyperparameter tuning and there are no independent samples left for testing. This is confirmed by the comparison of the classification accuracy of the two scenarios. As shown in Table 4, except LDA, all other methods tuned following Scenario B slightly outperform their counterparts tuned following Scenario A.

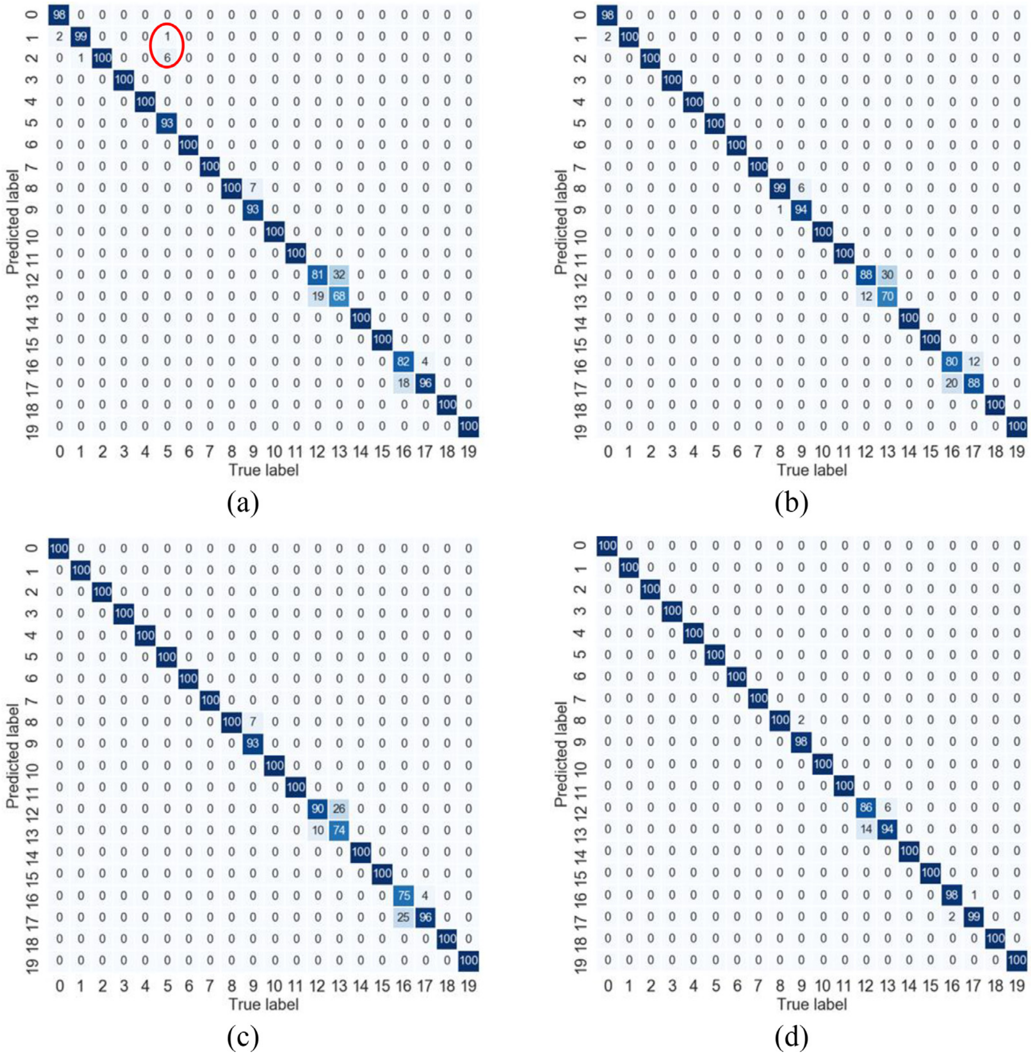


Fig. 14. Classification confusion matrices of different methods: (a) SVM, (b) ANN, (c) XGBoost, (d) Bagging with LDA as the base estimator. The confusion matrix of LDA is shown in Fig. 12 (b). The far-off misclassifications by SVM are circled by red ellipse.

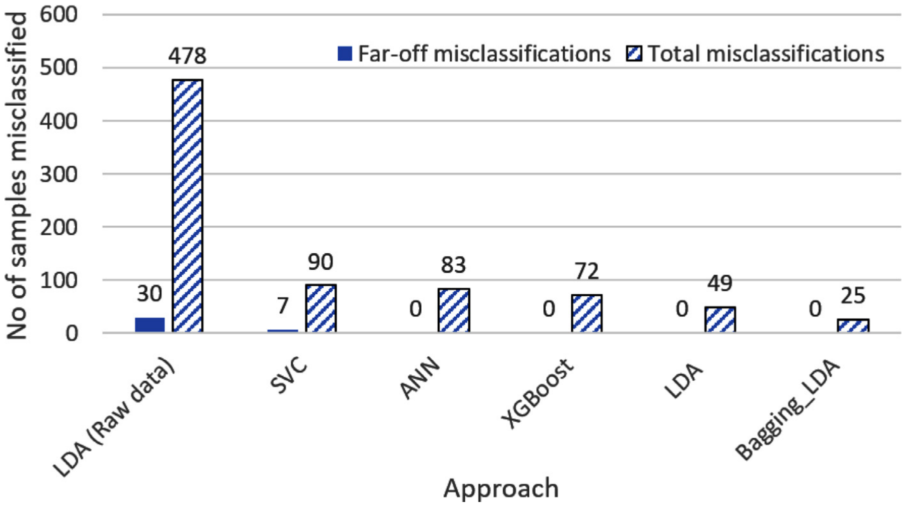


Fig. 15. Comparison of far-off misclassification and total misclassification of different approaches.

Table 4
Comparison of classification accuracy under two hyperparameter tuning scenarios.

Method	Scenario A	Scenario B
SVM	95.50 ± 3.79	96.40 ± 3.63
ANN	95.85 ± 4.15	96.35 ± 3.61
XGBoost	96.40 ± 3.70	96.40 ± 3.34
LDA	97.55 ± 2.89	97.55 ± 2.89
Bagging (LDA)	98.75 ± 2.29	99.35 ± 1.69

Therefore, the results reported previously in this work are all based on Scenario A for fair evaluation of all methods with independent test samples.

7. Conclusion

For the pulp and paper industry in the U.S., the pulping process has been identified as a major opportunity for improving energy efficiency and productivity. However, the implementation of model-based optimization, control and other advanced manufacturing technologies has been hindered by the lack of real-time sensing of woodchip MC under the harsh manufacturing environment. To overcome this bottleneck, we investigate the potential of an IIoT short-range Wi-Fi based woodchip MC sensing technology. The proposed technology takes the advantages of IIoT devices (e.g., toughness, connectivity, low-cost, small-size, etc.), while overcoming their shortcomings (e.g., the machine learning challenges of messy big data) by SPA-based feature engineering. Specifically, this work demonstrates that woodchip packing is a strong confounding factor to woodchip MC level, evidenced by its significant impact on both amplitude and phase of the collected CSI data. Although randomization is a good strategy to mitigate this confounding factor, it is not sufficient by itself. As a validation, we demonstrated that classification using raw CSI data results in not only low classification accuracy, but also many far-off misclassifications where the predicted MC class is off its true class by more than one level. The result also illustrates that classification accuracy alone is not a good performance metric, and the practical implications (e.g., cost) of misclassification must also be considered. We show that SPA-based feature engineering framework is a systematic approach for generating physically and statistically meaningful features compared to other kernel-type or algorithmically generated (*e.g.*, square, square root, exponential, etc.) features that are often unintuitive. Through simple feature selection such as PCA, the mean difference of consecutive subcarriers (MDCSs) of CSI amplitude were found to be robust features that are not only highly sensitive to MC levels but also highly insensitive to woodchip packing. Using MDCSs as features, we demonstrated the superior classification performance of using CSI data collected off all three antennas compared to that of using any single antenna. Finally, using MDCSs from all three antennas, we investigate the representative state-of-the-art classification techniques, including LDA, SVM, ANN and ensemble learning methods including bagging with LDA and gradient boosting with XGBoost. The results showed that LDA and its bagging extension perform the best among all methods, achieving overall classification accuracy of 98~99%. In addition, when MDCSs are used as features, only SVM results in far-off misclassifications, while all other methods only result in nearest-neighbor misclassifications, which is a significant improvement compared to when raw CSI data were used as features.

It is worth noting that although woodchip packing has significant impact to the collected CSI data (both amplitude and phase responses), its impact to MC classification is completely eliminated after we selected SPA features that are totally insensitive to packing. Although the randomization is done by shaking the same woodchips within a given volume - which means the volume density of the sample is about the same, the linear density (*i.e.*, linear void/packing fraction) actually varies significantly. If we assume linear paths of the Wi-Fi signal propagation, shuffling even the same woodchips can introduce significant variations to the linear void (or packing) fraction along the straight lines between the injector and the three receivers, as evidenced by the significant changes in the amplitude and phase responses of the CSI data. However, our results show that the selected SPA features (*i.e.*, mean difference of consecutive subcarriers of CSI amplitude) are insensitive to the shuffling, as evidenced by the high classification accuracy of independent (*i.e.*, differently shuffled) testing samples. Therefore, we can conclude that the selected SPA features are insensitive to the void fraction (or packing density) of the woodchips. This is particularly convincing when we consider the excellent performance of the technology at the low MC range where there is only 0.05% change in MC level but significant change in linear void fraction along the Wi-Fi propagation paths due to shuffling. Nevertheless, it is desirable to test woodchips with different sizes to further validate the technology. We envision that, when implemented to real industrial applications, some form of random sampling can be implemented to obtain multiple MC estimations, and some form of aggregation (*e.g.*, average) of different measurements can be used to obtain a reliable estimation of the MC level for a large quantity of woodchips.

It is also worth noting that this work only establishes the feasibility of this technology in the lab using a box. Whether the technology can be applied in more flexible settings, such as woodchips not in a box but in a pile on a fixed or moving surface (*e.g.*, a conveyor belt), requires further investigation. There is no doubt that the problem will be more challenging than what has been studied in this work, which is under a much better controlled environment in a lab. In addition, we only demonstrate the success of classification-based woodchip MC estimation in this work, while our preliminary results have shown that the regression based MC estimation is much more challenging for this application. This is due to the fact that, although MDCSs of CSI amplitude enables linear separation of different MC levels, the functional relationship between CSI data and woodchip MC values is actually much more complicated and research in this area is ongoing.

Author contribution

QH conceived the idea, supervised the experimental design and modeling work, led the writing and revision of the manuscript; KS participated in the experimental design and conducted the experiments, analyzed the results, and participated in the manuscript drafting and revision. KS and QH read and approved the final manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgement

Financial support from the National Science Foundation under grant number NSF-CBET #1805950 is gratefully acknowledged.

References

Aggarwal, C.C., 2018. Neural Networks and Deep Learning - A Textbook, Machine Learning. Springer.

Ahamed, I., Vijay, M., 2017. Comparison of different diversity techniques in MIMO antennas. In: 2017 2nd International Conference on Communication and Electronics Systems (ICCES). IEEE, pp. 47–50.

Amaral, E.A., Santos, L.M., Costa, E.V.S., Trugilho, P.F., Hein, P.R.G., 2020. Estimation of moisture in wood chips by near infrared spectroscopy. Maderas Cienc. y Tecnol. doi:10.4067/S0718-221X2020005000304.

ASTM, 2016. Standard test methods for moisture content of wood ASTM D4442. 1983. Annu. B. ASTM Stand. 431–445.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.

Bergstra, J., Yamins, D., Cox, D., 2013. Hyperopt: a python library for optimizing the hyperparameters of machine learning algorithms. In: Proceedings of the 12th Python in Science Conference, pp. 13–19. doi:10.25080/majora-8b375195-003.

Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. In: Advances in Neural Information Processing Systems, pp. 2546–2554.

Bishop, C.M., 2006. Machine Learning and Pattern Recognition, Information Science and Statistics. Springer, Heidelberg Springer, Heidelberg.

Breiman, L., 1999. Pasting small votes for classification in large databases and on-line. Mach. Learn. 36, 85–103. doi:10.1023/a:1007563306331.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. doi:10.1007/bf00058655.

Brueske, S., Kramer, C., Fisher, A., 2015. Bandwidth Study on Energy Use and Potential Energy Saving Opportunities in US Pulp and Paper Manufacturing. Energetics.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2, 121–167. doi:10.1023/A:1009715923555.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. doi:10.1145/2939672.2939785.

Choi, H., Kwon, J.S., 2019a. Modeling and control of cell wall thickness in batch delignification. Comput. Chem. Eng. 128, 512–523.

Choi, H., Kwon, J.S., 2019b. Multiscale modeling and control of Kappa number and porosity in a batch-type pulp digester. AIChE J. 65, e16589.

Choi, H., Son, S.H., Kwon, J.S., 2021. Inferential model predictive control of continuous pulping under grade transition. Ind. Eng. Chem. Res. 60, 3699–3710.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. doi:10.1023/A:1022627411411.

Couceiro, J., Lindgren, O., Hansson, L., Söderström, O., Sandberg, D., 2019. Real-time wood moisture-content determination using dual-energy X-ray computed tomography scanning. Wood Mater. Sci. Eng. doi:10.1080/17480272.2019.1650828.

Daassi-Gnaba, H., Oussar, Y., Merlan, M., Ditchi, T., Géron, E., Holé, S., 2018. Moisture content recognition for wood chips in pile using supervised classification. Wood Sci. Technol. 52, 1195–1211.

Daassi-Gnaba, H., Oussar, Y., Merlan, M., Ditchi, T., Géron, E., Holé, S., 2017. Wood moisture content prediction using feature selection techniques and a kernel method. Neurocomputing 237, 79–91.

Fridh, L., Eliasson, L., Bergström, D., 2018. Precision and accuracy in moisture content determination of wood fuel chips using a handheld electric capacitance moisture meter. Silva Fenn. 52.

Friedman, J., Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning, Elements. Springer series in statistics, New York doi:10.1007/978-0-387-84858-7.

Halperin, D., Hu, W., Sheth, A., Wetherall, D., 2012. 802.11 with multiple antennas for dummies. ACM SIGCOMM Comput. Commun. Rev. doi:10.1145/1672308.1672313.

Halperin, D., Hu, W., Sheth, A., Wetherall, D., 2011. Tool release: gathering 802.11n traces with channel state information. ACM SIGCOMM Comput. Commun. Rev..

He, Q.P., Wang, J., 2018a. Statistical process monitoring as a big data analytics tool for smart manufacturing. J. Process Control 67, 35–43. doi:10.1016/j.jprocont.2017.06.012.

He, Q.P., Wang, J., 2018b. Statistics pattern analysis: a statistical process monitoring tool for smart manufacturing. Comput. Aided Chem. Eng. 44, 2071–2076. doi:10.1016/B978-0-444-64241-7.50340-2.

He, Q.P., Wang, J., 2011. Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes. AIChE J. 57, 107–121. doi:10.1002/aic.12247.

He, Q.P., Wang, J., Gilicia, H.E., Stuber, J.D., Gill, B.S., 2012. Statistics Pattern Analysis Based Virtual Metrology for Plasma etch Processes. American Control Conference (ACC), pp. 4897–4902.

He, Q.P., Wang, J., Shah, D., 2019. Feature space monitoring for smart manufacturing via statistics pattern analysis. Comput. Chem. Eng. 126, 321–331.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844. doi:10.1109/34.709601.

Hu, J., Peng, H., Liu, T., Yao, X., Wu, H., Lu, P., 2019a. A flow sensing method of power spectrum based on piezoelectric effect and vortex-induced vibrations. Measurement 131, 473–481.

Hu, P., Yang, W., Wang, X., Mao, S., 2019b. MiFi: Device-free wheat mildew detection using off-the-shelf wifi devices. In: 2019 IEEE Global Communications Conference, GLOBECOM 2019 - Proceedings, pp. 1–6.

Hultnäs, M., Fernandez-Cano, V., 2012. Determination of the moisture content in wood chips of Scots pine and Norway spruce using Mantex desktop scanner based on dual energy X-ray absorptiometry. J. Wood Sci. doi:10.1007/s10086-012-1260-z.

Komer, B., Bergstra, J., Eliasmith, C., 2014. Hyperopt-Sklearn: automatic hyperparameter configuration for Scikit-learn. In: Proceedings of the 13th Python in Science Conference, pp. 32–37. doi:10.25080/majora-14bd3278-006.

Kramer, K.J., Masanet, E., Xu, T., Worrell, E., 2011. Energy efficiency improvement and cost saving opportunities for the pulp and paper industry. Improving Energy Efficiency and Greenhouse Gas Reduction in the Pulp and Paper Industry. Nova Science Publishers, Inc.

Liang, L., Fang, G., Deng, Y., Xiong, Z., Wu, T., 2019. Determination of moisture content and basic density of poplar wood chips under various moisture conditions by near-infrared spectroscopy. For. Sci..

Louppe, G., Geurts, P., 2012. Ensembles on random patches. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 346–361. doi:10.1007/978-3-642-33460-3_28.

Martin, N., Anglani, N., Einstein, D., Khrushch, M., Worrell, E., Price, L.K., 2000. Opportunities to improve energy efficiency and reduce greenhouse gas emissions in the US pulp and paper industry. Lawrence Berkeley Natl. Lab.

Merlan, M., Ditchi, T., Oussar, Y., Holé, S., Géron, E., Lucas, J., 2019. Resonant half-wave antenna for moisture content assessment in wood chips. Meas. Sci. Technol..

Nielsen, M., 2016. Neural Networks and Deep Learning, Neural Networks and Deep Learning. Determination Press.

Pan, P., McDonald, T., Fulton, J., Via, B., Hung, J., 2017. Simultaneous moisture content and mass flow measurements in wood chip flows using coupled dielectric and impact sensors. Sensors 17, 20.

Pan, P., McDonald, T.P., Via, B.K., Fulton, J.P., Hung, J.Y., 2016. Predicting moisture content of chipped pine samples with a multi-electrode capacitance sensor. Biosyst. Eng. doi:10.1016/j.biosystemseng.2015.12.005.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res..

Rahman, M., Avelin, A., Kyprianidis, K., 2020. A review on the modeling, control and diagnostics of continuous pulp digesters. Processes 8, 1–26. doi:10.3390/pr8101231.

Reeb, J., Milota, M., 1999. Moisture content by the oven-dry method for industrial testing. WDKA 66–74. doi:10.1016/j.apsusc.2013.11.072.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks Cambridge.

Shah, D., Hancock, A., Skjellum, A., Wang, J., He, Q.P., 2017. Challenges and opportunities for IoT-enabled cybermanufacturing: what we learned from an IoT-enabled manufacturing technology testbed. In: Proceedings of Foundations of Computer Aided Process Operations /Chemical Process Control, p. 66.

Shah, D., Wang, J., He, Q.P., 2020. Feature engineering in big data analytics for iot-enabled smart manufacturing–comparison between deep learning and statistical learning. Comput. Chem. Eng., 106970.

Shah, D., Wang, J., He, Q.P., 2019a. An internet-of-things enabled smart manufacturing testbed. In: IFAC-PapersOnLine. Florianopolis, Brazil, pp. 562–567. doi:10.1016/j.ifacol.2019.06.122.

Shah, D., Wang, J., He, Q.P., 2019b. A feature-based soft sensor for spectroscopic data analysis. J. Process. Control 78, 98–107.

Suthar, K., Shah, D., Wang, J., He, Q.P., 2019. Next-generation virtual metrology for semiconductor manufacturing: a feature-based framework. Comput. Chem. Eng. 127, 140–149. doi:10.1016/j.compchemeng.2019.05.016.

Suthar, K., Shah, D., Wang, J., Peter He, Q., 2018. Feature-based virtual metrology for semiconductor manufacturing. Computer Aided Chemical Engineering doi:10.1016/B978-0-444-64241-7.50342-6.

Theodoridis, S., 2015. Neural networks and deep learning. Machine Learning. Academic Press doi:10.1016/b978-0-12-801522-3.00018-5.

Wang, J., He, Q.P., 2010. Multivariate statistical process monitoring based on statistics pattern analysis. Ind. Eng. Chem. Res. 49, 7858–7869. doi:10.1021/ie901911p.

Xie, Y., Li, Z., Li, M., 2018. Precise power delay profiling with commodity Wi-Fi. IEEE Trans. Mob. Comput. 18, 1342–1355.

Yang, W., Wang, X., Cao, S., Wang, H., Mao, S., 2018a. Multi-class wheat moisture detection with 5GHz Wi-Fi: a deep LSTM approach. In: 2018 27th International Conference on Computer Communication and Networks (ICCCN). IEEE, pp. 1–9.

Yang, W., Wang, X., Song, A., Mao, S., 2018b. Wi-wheat: contact-free wheat moisture detection with commodity WiFi. In: 2018 IEEE International Conference on Communications (ICC). IEEE, pp. 1–6.