2020 American Control Conference
Denver, CO, USA, July 1-3, 2020

Using Channel State Information for Estimating Moisture Content in Woodchips via 5 GHz Wi-Fi*

Kerul Suthar, Jin Wang, *Member, IEEE*, Zhihua Jiang, Q. Peter He, *Senior Member, IEEE*

Abstract— **For the pulping process in a pulp & paper plant that uses wood as a raw material, it is important to have real-time knowledge about the moisture content of the woodchips so that the process can be optimized and/or controlled correspondingly to achieve satisfactory product quality while minimizing the consumption of energy and chemicals. Both destructive and non-destructive methods have been developed for estimating moisture content in woodchips, but these methods are often lab-based that cannot be implemented online, or too fragile to stand the harsh manufacturing environment. To address these limitations, we propose a non-destructive and economic approach based on 5 GHz Wi-Fi and use channel state information (CSI) to estimate the moisture content in woodchips. In addition, we propose to use statistics pattern analysis (SPA) to extract features from raw CSI data of amplitude and phase difference. The extracted features are then used for classification model building using linear discriminant analysis (LDA) and subspace discriminant (SD) classification. The woodchip moisture classification results are validated using the oven drying method.**

I. Introduction

The pulp and paper industry is the third largest consumer of energy in the US industrial sector, so it has tremendous opportunities to improve its energy efficiency and productivity. The pulping process, which converts woodchips into pulp by displacing lignin from cellulose fibers, is one of the most important operations in a pulp and paper mill. Because the pulping process uses wood as a raw material, it is important to have real-time knowledge about the moisture content in the woodchips so that the process can be optimized and/or controlled correspondingly to achieve satisfactory product quality while minimizing the consumption of energy and chemicals. 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 is a major source of disturbance that affects the cooking performance, as it dilutes the white liquor concentration therefore reducing the delignification reaction rate. Currently, the woodchip moisture content is not measured in real-time due to the lack of affordable, reliable and easy-to-maintain sensors; instead, woodchip moisture content is commonly measured only four time per year corresponding to the four seasons, and used to determine the operation parameters such as chemical usage. 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.

To address the above-mentioned challenges, we propose a non-destructive and economic approach based on 5 GHz Wi-Fi and use channel state information (CSI) to predict the moisture content in woodchips. We extract CSI by modifying the open source device drivers for Intel Wi-Fi link 5300 network interface card (IWL5300 NIC) based on CSITool [1]. CSI contains information about the channel in the form of individual data subcarriers capturing indoor channel features such as the effect of scattering, fading and power decay with distance. Modern Wi-Fi systems are equipped with orthogonal frequency division multiplexing (OFDM), dividing the data into multiple orthogonal subcarrier groups which solves the issue of selective frequency fading [2]. CSI has been used for indoor localization, device-free sensing including fall detection, activity recognition, and heart rate monitoring [3]. In addition, CSI and phase difference data have been successfully used for moisture detection in wheat [3].

In this work, we collect CSI and phase difference data using IWL5300 NIC by configuring the transmitter and receiver in injection and monitor mode respectively. We use Lenovo ThinkPad systems equipped with Linux based OS 14.02 and kernel version 4.2 due to the version-specific selectivity of CSI tool. Both systems are equipped with IWL5300 NIC with a modified driver and firmware for data collection. Our work includes CSI data collection, preprocessing, outlier detection, offline training, and online testing. We collect CSI amplitude and phase difference for 20 different moisture levels ranging from 11% to 53%. We experimentally validate the feasibility of using CSI amplitude and phase difference data to estimate moisture in woodchips. Compared to wheat moisture detection [3], woodchip moisture estimation is much more challenging due to the much big and more heterogeneous size of the woodchips. Because of that, the woodchip arrangement in the container is expected to have significant impact on CSI data. This effect must be excluded or filtered out from the model so that consistent moisture estimation can be obtained. The main contribution of this work is to develop a robust model that is insensitive to the shuffling

*Research supported by NSF-CBET # 1805950.

All authors are with the Department of Chemical Engineering, Auburn University, Auburn, AL 36830 USA

Corresponding author: Q. Peter He phone: 334-844-7602; fax: 334-844-2063; e-mail: qhe@auburn.edu.

978-1-5386-8266-1/$31.00 ©2020 AACC

of the woodchips. To achieve this goal, the statistics pattern analysis (SPA) framework that we developed previously [4]–[6] is adopted in this work to build a multivariate statistical model based on different classification approaches for multi-class moisture detection in woodchips.

The remainder of the paper is organized as follows. A brief background on CSI and feasibility study are presented in Section II. Section III outlines the experimental setup and data collection. The classification model development is discussed in Section IV. The results are discussed in Section V followed by conclusions in Section VI.

II. Channel State Information and Feasibility Study

A. Channel State Information (CSI)

Using Wi-Fi cards such as IWL5300, it is convenient to collect CSI measurements that record the channel variation during propagation of wireless signals. After being transmitted from a source, the wireless signal is expected to experience impairments caused by obstacles before the signal reaches the receiver. CSI can reflect indoor channel characteristics such as multipath effect [7], shadowing, fading, and delay. In comparison to the received signal strength (RSS), CSI amplitude and phase difference data are relatively stable. Orthogonal frequency-division multiplexing (OFDM) is a method of digital signal modulation where a single data stream is split into multiple orthogonal subcarrier at different frequencies to avoid interference and crosstalk. IWL5300 NIC implements an OFDM system with 56 subcarriers. We are able to read information for 30 out of the 56 subcarriers using the CSItool, which is built on IWL5300 NIC using a custom modified firmware and open source Linux wireless drivers [1]. IWL5300 NIC provides 802.11n CSI for 30 out of the 56 subcarriers, which is about one group for every 2 subcarriers at 20 MHz or one in 4 at 40 MHz. 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. The Channel response of the i^{th} subcarrier can be given as:

$$CSI_i = |CSI_i| exp\{\angle CSI_i\} \tag{1}$$

where $|CSI_i|$ is the amplitude response of the i^{th} subcarrier, $\angle CSI_i$ the phase response and exp the exponential function. The three antennas of the IWL5300 NIC have different CSI features, which can be exploited to improve the diversity of training and test samples as show in Fig. 1.

B. Feasibility Test

With the help of CSItool, CSI data for 30 out of 56 subcarriers from IWL5300 NIC can be obtained for each packet. Two laptops equipped with IWL5300 NIC and modified drivers with specific Linux kernels are used to collect CSI data. One of the devices is set in injection mode while the other is set in monitor mode to collect 5 GHz CSI amplitude, phase and phase difference. One antenna is used on the transmitter side, while three antennas are used on the transmitter side to take the advantage of the (multiple-input multiple-output) MIMO systems for improving the diversity of the training and test samples [8]. In this work, it is ideal 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 are great for point-to-point connection which would also mean that unwanted interferences from other sources are expected to be minimized. We use panel antennas ALFA (ALFA Network, Taiwan) with 66° horizontal beam-width and 16° vertical beam-width.

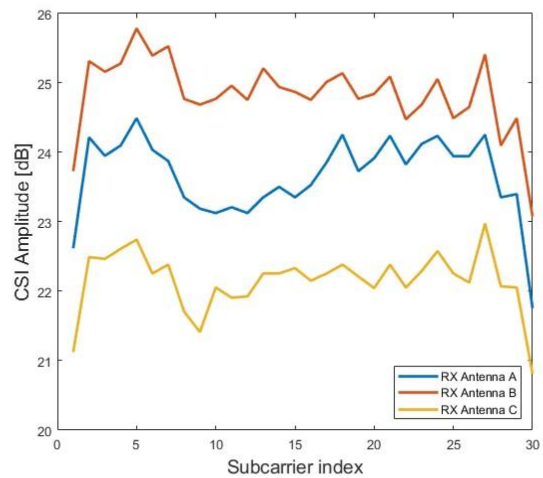


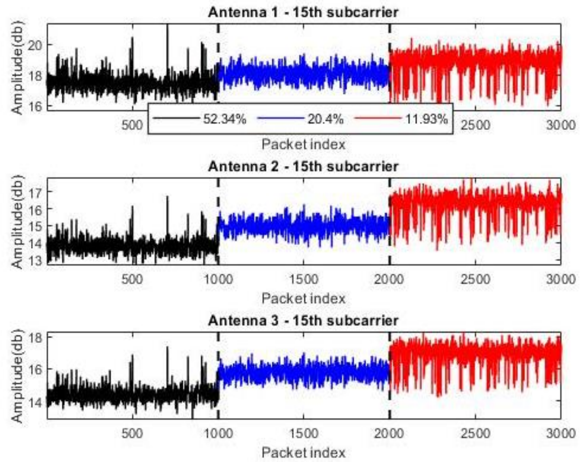
Figure 1 CSI amplitude for all the three antennas of IWL5300 NIC over 30 subcarriers

To establish the experimental feasibility of CSI being able to separate woodchips based on moisture levels, we collect CSI for 3 different moisture levels, i.e., 52.34%, 20.40% and 11.93%. Fig. 2 shows the CSI amplitude and phase differences for all three different moisture levels for the 15^{th} subcarrier. As shown in Fig. 2, there are distinctive differences in both amplitude and phase difference of different moisture levels from all three antennas.

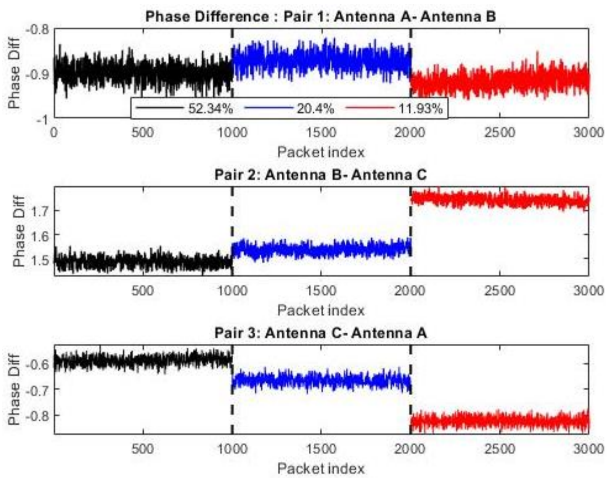
III. Experimental Setup and Data Collection

A. Experimental setup

With the results from the feasibility test in Section II, we design an experimental setup with antennas positions fixed on an acrylic sheet. We use commodity laptops and Wi-Fi to implement the proposed system architecture. Two Lenovo T400s systems are equipped with IWL5300 NIC along with their modified device driver for CSI data accessibility. We set one of the systems in transmitter mode while the other in receiver mode and inject packets from the transmitter using one antenna. We use three antennas on the receiver side to explore the variability of signals received at all the three receiving antennas and collect CSI data for each received packet. The laptops run 32-bit Ubuntu Linux 14.04 with kernel version 4.2.027, as CSItool is compatible with kernel version 3.2-4.2. The transmitter and receiver are separated at a distance



(a)



(b)

Figure 2 CSI amplitude (a) and phase difference (b) from the three antennas at three different moisture levels

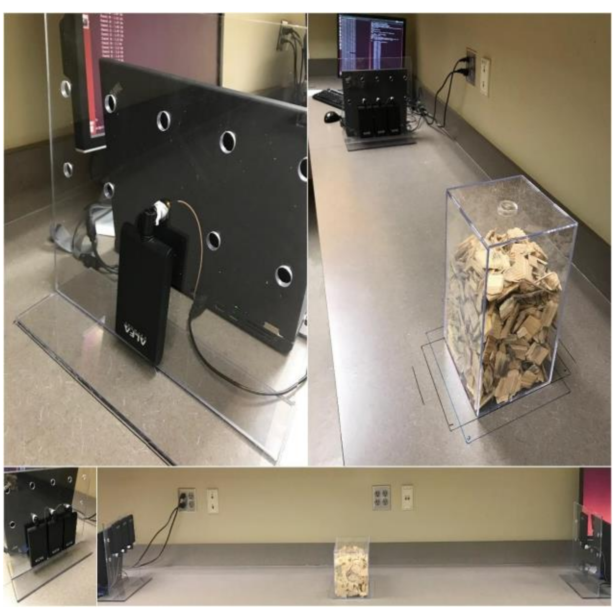


Figure 3 Experimental setup for CSI data collection

of 3m to explore the multipath effect [7]. The experimental setup is shown in Fig. 3. The woodchips at different moisture levels are placed in an acrylic container with an air-tight lid to avoid any changes in moisture while the data is being collected. Oven drying method [9][10] is used for measuring the actual moisture content of the woodchips.

B. Data collection

Data are collected for 20 different moisture levels ranging from 53.38 % to 11.81% on the wet basis. Oven drying was performed in the end to determine the oven dry weight, which is used to determine the actual moisture contents. As discussed previously, the woodchip arrangement in the container is expected to have significant impact on CSI data. To fully excite the system on the effect of shuffling, the woodchips within the airtight box are shuffled 10 times for each moisture level. In other words, for each moisture level, 10 datasets (i.e., samples) are collected corresponding to 10 shuffles. Therefore, there are totally 200 samples. For each sample, 1,000 packets were sent from the transmitter setup in injection mode to the three receiver antennas on the receiver side setup in monitoring mode. Ultimately, an efficient model should be able to estimate the correct moisture level of all the 10 shuffled samples that are at the same moisture level. The frequency during data-collection over a period of 8 days is the same and set at channel 64, i.e., 5.32 GHz to avoid any discrepancies in the data. The data collected includes moisture levels very closely separated both at a higher moisture level and at a lower moisture level to test out the efficiency of models. The 20 different moisture levels are plotted in Fig. 4. 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. Non-LOS (NLOS) will be investigated as a part of future work.

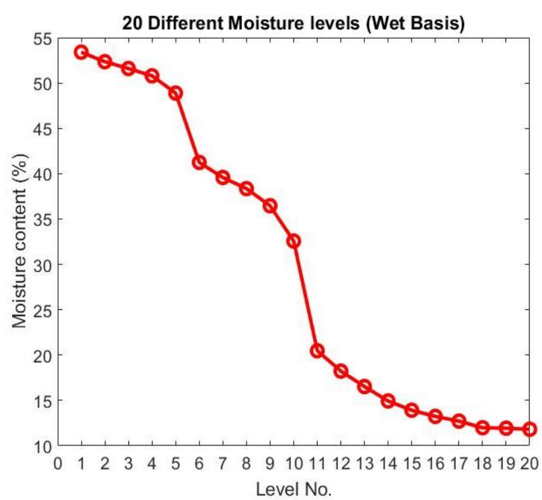


Figure 4 20 different moisture levels of woodchips

IV. MODEL BUILDING AND CLASSIFICATION

In this work, statistics pattern analysis (SPA) is utilized for feature generation. Instead of building relationships between the CSI amplitude/phase difference and moisture content, relationships are extracted between the statistics of the CSI amplitude/phase difference and woodchip moisture content.

A. Statistics pattern analysis-based moisture classification

In statistics pattern analysis (SPA), various statistics are used to quantify process characteristics, and instead of process variables themselves, statistics are used for modelling. SPA has been applied for fault detection [4][5], fault diagnosis [11], and virtual metrology or soft sensor [12]–[14].

The schematic diagram of SPA is shown in Fig. 5. In the first step, various statistics are extracted from the process variables, i.e., CSI data in our case.

$$\mathbb{P}: \boldsymbol{X} \rightarrow \boldsymbol{F} \tag{2}$$

where $\mathbb{P}$ denotes the operator that maps the 3D process data array $\boldsymbol{X} \in \mathcal{R}^{N\times S\times K}$ containing N samples, S amplitudes and phase differences of all subcarriers from K packets into a feature matrix $\boldsymbol{F} \in \mathcal{R}^{N\times S}$ 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. Other between-variable statistics, such as cross-correlations, are calculated similarly. In Fig. 5, $\boldsymbol{Y} \in \mathcal{R}^{N\times 1}$ denotes the moisture levels for N samples. In the second step, a regression or classification method can be used to extract relationships between the features and the response, i.e. moisture levels. 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.

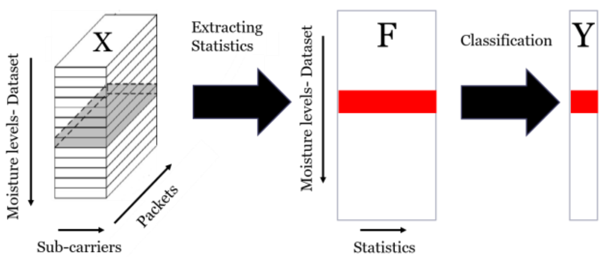


Figure 5 Schematic of SPA for classification

B. Model building

CSI data from woodchips are collected at 20 different moisture levels with 10 shuffles at each level, which results in $N = 200$ samples. For each sample, 1,000 packets are sent from the transmitter to the receivers. The complex value CSI matrix is decomposed into its respective amplitude $\boldsymbol{A}$ and phase $\boldsymbol{P}$. Phase difference from each antenna pair is calculated. The main reason for using phase difference instead of the phase itself is that the phase difference is relatively more stable in comparison to the phase itself [15].

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1j} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{ij} \end{bmatrix} \tag{3}$$

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & \cdots & p_{1j} \\ \vdots & \ddots & \vdots \\ p_{i1} & \cdots & p_{ij} \end{bmatrix} \tag{4}$$

where a_{ij} and p_{ij} denote the CSI amplitude and phase of signal from subcarrier j of packet i, respectively. If the raw CSI data to be used, the amplitude and phase difference matrices are to be vectorized, which results in K = 180,000 variables if both amplitude and phase differences from all three antennas and 30 subcarriers are used. Clearly, there are significantly more variables than samples. To reduce the dimension of variables, SPA is applied to extract statistics following the procedure outlined in the previous section, which results in S = 360 features when mean and standard deviations of both amplitude and phase difference from all three antennas and 30 subcarriers are used. In this work, data from all the 30 subcarriers are used and no subcarrier or variable selection is performed. However, subcarrier or variable selection is one of the future extensions of this work. For training, 8 samples are randomly selected from 10 shuffled samples at the same moisture level for each of the 20 different moisture levels, which results in 160 training samples containing half of the data covering all moisture levels. The rest 40 samples are used for testing after the classification model is trained to evaluate the performance of the trained model.

Next, classification is applied to build a relationship between the SPA features and moisture levels. As a part of preliminary analysis, two classification methods are compared in this work: linear discriminant analysis (LDA) and subspace discriminant (SD) classification based on random subspace algorithm [16]. For both methods, 20% of the data is held out for testing.

The Monte Carlo validation and testing (MCVT) procedure [14] is followed to repeat the random training/testing sample selection and model building/testing procedure and the average of 100 such MC runs are considered to evaluate the performance of the models. The accuracy of the models is evaluated based on how many moisture levels/classes are correctly classified on the test samples.

V. Results

In this work, we compare prediction accuracies for three different CSI scenarios where the models are built using 1) amplitude for all three antennas, 2) phase difference among all three antenna pairs and 3) amplitude and phase difference of all the three antennas. In this work, only mean and standard deviation of each variable are considered as these statistics capture the general behavior of the predictors. More statistics, especially higher order statistics (HOS) and statistics that capture between-variable relations such as cross-correlation, will be studied in our future work. In addition, as mentioned before, CSI data for 20 different moisture levels have been collected. In this work, we compare two different moisture scenarios: 1) model built using 10 different moisture levels, i.e., every other moisture level out of 20 moisture levels; and 2) A model built using all 20 moisture levels. Fig. 6 shows the comparison of the two classification methods for models built

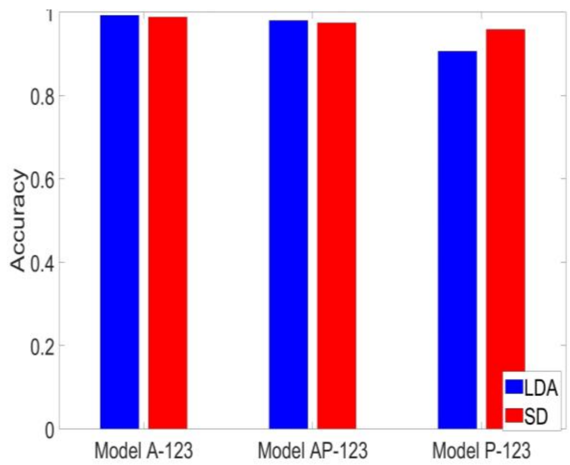


Figure 6 Classification accuracy using10 moisture levels

using data for 10 different moisture levels. Three different modelling scenarios compared are: Model A-123, where the amplitudes of all the three antennas are used; Model P-123, where the phase differences of all three antenna pairs are used; and finally Model AP-123 where both amplitudes and phase differences for all the three antennas are used. It can be seen that for 10 different moisture levels, both methods have a classification accuracy as high as 99%. Using only phase difference, accuracy is relatively lower, but still, the classification accuracy based on SD is 95.9%.

Fig. 7 shows the comparison similar to Fig. 6 when the model is trained and tested using all the moisture levels. The results show that as more moisture levels are included in the model, the classification accuracy decreases, which is expected. Nevertheless, the classification accuracy of SD using amplitudes from all three antennas is still high: 97.7%. It is worth noting that when a moisture level is misclassified, it is always misclassified as a moisture level not far from its true level. Fig. 8 shows the confusion matrix from one MC run of SD using 20 moisture levels. As can be seen from the figure, the misclassified samples are incorrectly classified as their respective neighboring classes. This is important for real applications as it means that even when the misclassification occurs, we can still obtain a reasonably good estimate of the

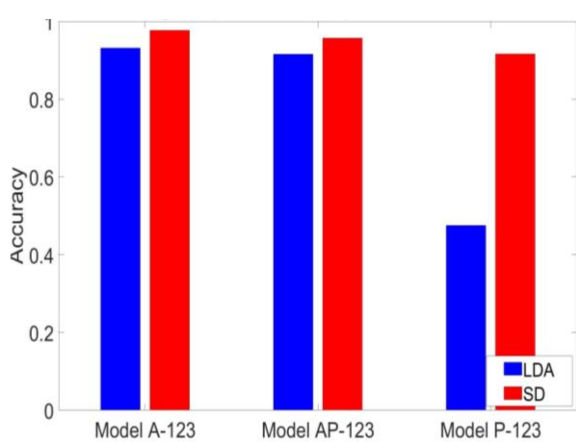


Figure 7 Classification accuracy using 20 moisture levels

moisture level. In addition, both Fig. 6 and 7 indicate that the features based on phase differences do not help with moisture classification, either used alone, or combined with features based on amplitudes. This result is different from [3] where phase difference was found to be more effective for wheat moisture classification. This difference may be due to the difference in feature extraction and/or the fact that woodchips have bigger size and are more heterogenous in both size and shape, which lead to greater variability in phase differences than that in amplitudes. Finally, SD classification performs similarly to LDA in the case with 10 moisture levels. But SD performs noticeably better than LDA in the case of 20 moisture levels. This is probably due to the fact that LDA often suffers from the small sample size with high dimensional features, while SD can address this problem by random sampling on features [16].

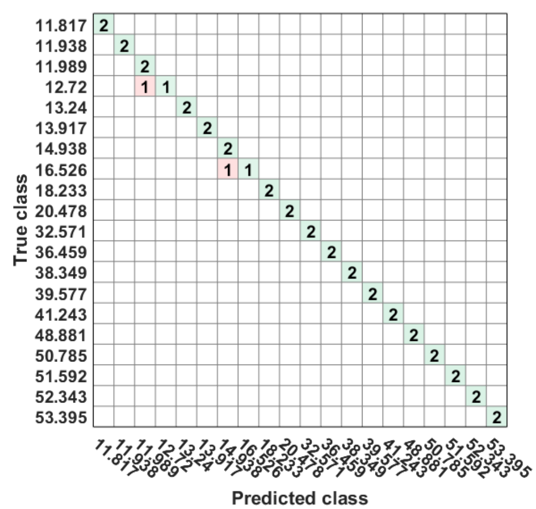


Figure 8 Confusion matrix of one MC run of SD on 20 moisture levels

Fig. 9 shows the comparison between 10 moisture levels and 20 moisture levels for SD. We can see that classification for 10 moisture levels yields better results than 20 moisture levels. This is expected as more moisture levels are introduced, the classification becomes more challenging.

Fig. 10 shows the importance of using data from all three antennas instead of just one antenna for 20 different moisture levels. The classification accuracy improves when features based on amplitudes from all three antennas are used. Similar behavior is observed for phase difference as well.

VI. Conclusion and Future work

In this work, we propose to use a non-destructive and economic approach based on 5 GHz Wi-Fi and to use collected channel state information (CSI) to estimate the moisture content in woodchips. Experiments were conducted to collect 200 samples of CSI data with 20 moisture levels. In addition, we propose to use statistics pattern analysis (SPA) to extract features from raw CSI data of amplitude and phase difference. The extracted features are then used for classification model building. In this work, we compare two classification methods using SPA features: linear

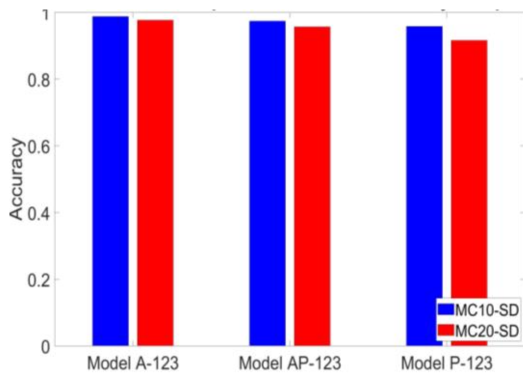


Figure 9 Classification accuracy of SD for 10 and 20 moisture levels

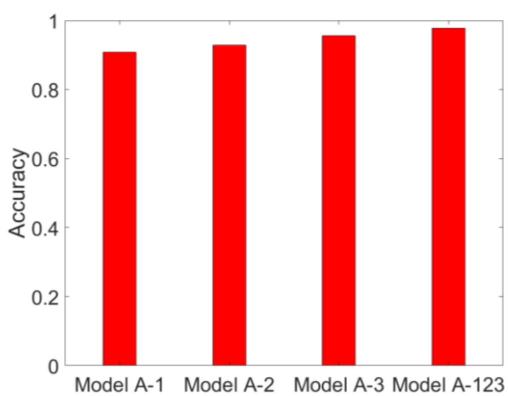


Figure 10 Importance of using all 3 antennas

discriminant analysis (LDA) and subspace discriminant (SD) classification. In general, SD classification performs better than LDA. The classification accuracy of SD for 10 moisture levels is as high as 98.8% and the accuracy for 20 moisture levels is as high as 97.7%. In both cases, the best performance is achieved when features based on amplitudes from all three antennas are used. Features based on phase differences do not help with moisture classification, either used alone, or combined with features based on amplitudes. This result is different from literature where phase difference was found to be more effective for wheat moisture classification. This difference may be due to the difference in feature extraction and/or the fact that woodchips have bigger size and are more heterogenous in both size and shape, which lead to greater variability in phase differences than that in amplitudes.

For future work, more statistics, especially higher order statistics (HOS) and statistics that capture between-variable relations such as cross-correlation, will be investigated. Feature selection will also be investigated to reduce the feature space dimension and to improve classification performance. Other classification techniques such as support vector machines (SVM) and artificial neural networks (ANN) will also be investigated.

REFERENCES

[1] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool Release: Gathering 802.11n Traces with Channel State Information," *ACM SIGCOMM Comput. Commun. Rev.*, 2011.

[2] X. Wang, S. Mao, S. Pandey, and P. Agrawal, "CA2T: Cooperative Antenna Arrays Technique for pinpoint indoor localization," in *Procedia Computer Science*, 2014.

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

[4] J. Wang and Q. P. He, "Multivariate Statistical Process Monitoring Based on Statistics Pattern Analysis," *Ind. Eng. Chem. Res.*, vol. 49, no. 17, pp. 7858–7869, Mar. 2010.

[5] Q. P. He and J. Wang, "Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes," *AIChE J.*, vol. 57, no. 1, pp. 107–121, Jan. 2011.

[6] Q. P. He and J. Wang, "Statistical process monitoring as a big data analytics tool for smart manufacturing," *J. Process Control*, vol. 67, pp. 35–43, 2018.

[7] C. Used and C. Study, "Multipath and Diversity," pp. 1–7.

[8] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "802.11 with multiple antennas for dummies," *ACM SIGCOMM Comput. Commun. Rev.*, 2012.

[9] ASTM, "Standard test methods for moisture content of wood ASTM D4442," *1983 Annu. B. ASTM Stand.*, no. November, pp. 431–445, 2016.

[10] J. Reeb and M. Milota, "MOISTURE CONTENT BY THE OVEN-DRY METHOD FOR INDUSTRIAL TESTING Weight of water Weight of wood MC -," *Wdka*, pp. 66–74, 1999.

[11] H. Galiciaa, Q. Heb, and J. Wanga, "Statistics Pattern Analysis based fault detection and diagnosis," *CPC VIII Conf.*, 2012.

[12] Q. P. He, J. Wang, H. E. Gilicia, J. D. Stuber, and B. S. Gill, "Statistics pattern analysis based virtual metrology for plasma etch processes," in *American Control Conference (ACC), 2012*, 2012, pp. 4897–4902.

[13] K. Suthar, D. Shah, J. Wang, and Q. P. He, "Next-generation virtual metrology for semiconductor manufacturing: A feature-based framework," *Comput. Chem. Eng.*, vol. 127, pp. 140–149, 2019.

[14] D. Shah, J. Wang, and Q. P. He, "A feature-based soft sensor for spectroscopic data analysis," *J. Process Control*, vol. 78, pp. 98–107, 2019.

[15] X. Wang, C. Yang, and S. Mao, "A TensorBeat: Tensor Decomposition for Monitoring Multi-Person Breathing Beats with Commodity WiFi," vol. V, pp. 1–28, 2016.

[16] X. Zhang and Y. Jia, "A linear discriminant analysis framework based on random subspace for face recognition," *Pattern Recognit.*, vol. 40, no. 9, pp. 2585–2591, 2007.