

# Machine Learning -based Optimization of Biomass Drying Process: Application of Utilizing Data Center Excess Heat

Henna Tiensuu\* Virpi Leinonen\*\* Jani Isokääntä\*\*  
Jaakko Suutala\*

\* *Biomimetics and Intelligent Systems Group, University of Oulu, P.O. BOX 4500, FI-90014, Oulu, Finland (e-mail: firstname.lastname@oulu.fi).*

\*\* *SFTec Oy, Oulu (e-mail: jani.isokaanta@sftec.fi)*

---

**Abstract:** This research explores the feasibility of using excess heat from data centers for biomass drying, enhancing the biomass energy value. A predictive model was developed to estimate exhaust air humidity from the dryer, indirectly indicating biomass moisture. Machine learning techniques, including linear regression model (LM), gradient boosting machines (GBM), eXtreme gradient boosting (XGBoost), random forest (RF), and multilayer perceptron (MLP), were used. Tree-based models GBM, RF, and XGBoost achieved a coefficient of determination ( $R^2$ ) of 0.88–0.89. Methods were enhanced with transparency through explainable artificial intelligence (XAI) techniques, which facilitated the analysis and visualization of humidity fluctuations. Key factors affecting drying efficiency include weather conditions, supply air humidity, and fan speed. The study provides actionable insights for optimizing the drying process, improving system air tightness, and advancing sustainable energy utilization through AI-driven solutions. The developed model enables future dynamic control of drying processes.

*Keywords:* process monitoring, explainable AI, predictive modeling, decision support, data centers

---

## 1. INTRODUCTION

Excess heat utilization from data center is widely investigated and is high interest of service providers as the climate targets, questions and demands are also increasing around the data center field. The utilization of data center excess heat especially in Nordics is concentrated to utilization in district heating (Wahlroos et al., 2018), as it is a well-established heating method in the area. District heating production in general still relies on fossil fuels and for example in Finland largely also to wood-based bioenergy. The use of wood-based bioenergy is increasing because of climate actions and increased price of CO<sub>2</sub> allowances. For energy production the wood based raw materials like wood chips are highly utilized. One option to utilize the waste heat of data center could be drying of biomass. As very even quality continuous heat flow is available from the datacenter and the drying of biomass as such requires lots of energy, but at the same time the energy value of the dried material increase by drying. The biomass drying has been considered earlier an option for waste utilization as an external process (Wahlroos et al., 2018) and now it was tested in the industrial scale.

In industrial settings, the challenges posed by demanding measurement conditions can significantly impact on quality of the data, and the data can come from several different sources. This multi-source data, often varied in format and structure, adds complexity to the analysis process. Despite these challenges, machine learning (ML)

methods are widely employed to diagnose, optimize, and enhance the quality and efficiency of complex manufacturing processes. However, the volume of data generated in industrial environments can be immense, further complicating the analysis process. ML techniques offer a means to extract valuable insights from these large and diverse datasets, enabling the prediction of process outcomes and the identification of relationships between different process parameters (He et al., 2009).

Previously, wood drying processes have been successfully optimized using ML methods across various applications. Ascher et al. (2022) have shown that ML methods have great potential towards modelling the biomass and waste gasification and pyrolysis processes and predicting the processes' product yields and properties. Chai et al. (2019) utilized feed forward neural network to simulate wood moisture content during the high-frequency drying, while Onsree and Tippayawong (2021) achieved accurate predictions of solid products yields from biomass torrefaction processes using gradient boosting machines (GBM). Studies have explored traditional physical methods in this area as well. For example, Li et al. (2012) investigated integrating a drying process into a power generation plant, using waste energy from process industries. These sources included low-grade heat, such as flue gas or hot cooling water for superheated steam. Additionally, Gebreegziabher et al. (2013) developed a physical model to determine the optimum drying level of wood chips. Furthermore, Li et al. (2022) examined using steel heat carriers for waste heat re-

covery and drying of high-moisture biomass in direct-fired power generation, achieving a 77.4% thermal efficiency in waste heat recovery and reducing fuel moisture content.

To fully leverage data-driven process modeling, the model results must be transparent to humans managing the manufacturing process. Hence, employing explainable machine learning methods is advisable (Goebel et al., 2018). Transparency should be increased, particularly when the model structure fails to explain the root causes behind the outcomes (Hagrass, 2018). This transparency not only aids in understanding but also facilitates the optimization of the process, enabling humans to make informed decisions based on the clear explanations.

In this research, first, a simple linear regression model is trained to predict biomass moisture after the drying process using a small dataset. Then, a model for predicting the absolute humidity of the air exhausted from the dryer is developed using LM (James et al., 2013), eXtreme gradient boosting (XGBoost) (Chen and Guestrin, 2016), GBM (Hastie et al., 2001), random forest (RF) (Breiman, 2001), and multilayer perceptron (MLP) (Rosenblatt, 1958; Bishop, 1995) modeling methods, and the root causes behind the undesirable humidity levels are identified. The modeling results have been analysed and visualized using XAI methods.

This study represents a pioneering effort in the field, to our knowledge, being the first to utilize waste heat from data centers for biomass drying and optimize the process through exhaust air humidity prediction. The article is organized as follows: In Section 2, the biomass drying process is described. Section 3 explains the data collection process. The modeling and visualization methods are introduced in Section 4, while Section 5 presents model training and results. Finally, the discussion and conclusion are found in Sections 6 and 7, respectively.

## 2. BIOMASS DRYING PROCESS

The biomass drying using waste heat from a small data center, the Boden Type Data Center (BTDC) located in Boden, Sweden, was tested by installing an industrial scale dryer unit, ModHeat®, at the data center site. The biomass drying test setup is depicted in Fig. 1. The drying air, in this case waste heat from data center was taken straight from the BTDC (from area of 60 m<sup>2</sup>) to the drier. The dryer's fan was used to suck the warm air from the data center to the dryer. The dried test material was a normal energy wood chip.

The wet wood chips were fed to the hopper where the material was then fed by conveyer belt to the dryer's materials feeder. From the material feeder the biomass was fed to dryer. Inside the dryer material was circulated from drying level to another, as the dryer consist of five drying levels. The drying air from the data center was in contact with the material inside the dryer and was circulated in the opposite direction with the material flow. The drying air was circulated by the exhaust fan of the dryer and the moist exhaust air was directed to atmosphere after drying. The dried material discharged from the bottom level of the dryer to the conveyer and unloaded to the skip.

The main feature of the drying test campaign at the data center end was the hot aisle temperature (30/42°C) representing a traditional and high-performance computing data center, which was achieved by the varying the operation modes for the cooling equipment. At the dryer end, the main features were the air flow rate at inlet (50/100%), which was adjusted with the dryer's fan speed and the material feeding rate (50/75/100 %), which was adjusted to control the material flow volume to the dryer. The material flow rates were 2, 2.7, and 3 m<sup>3</sup>/h.



Fig. 1. The layout of the biomass drying test setup.

## 3. DATA COLLECTION

The test campaign lasted about two weeks from September 14<sup>th</sup> to 24<sup>th</sup>, 2020, during which the test setup was instrumented for a data collection. The data collection consists of data from four different systems, of which three were inputs. One of the systems was the internal system of the data center (provided by EcoCooling), which collected environmental data from the computer room evaporation units (CREC). The CREC handles the server cooling and kept the data center aisles at target temperatures. The second data collection system was the ModHeat® dryer data collection system gathering the information about temperature, relative humidity, and air speed going in and out of the dryer. The third system (provided by RISE Research Institutes of Sweden) gathered the data about temperature, relative humidity, air volume flows entering and exhausting the dryer. As the inlet and outlet temperatures and relative humidity of air were the main data points for dryer's efficiency evaluation both data loggers were used to avoid data loss. The fourth system collected the data related to electric power usage data of the servers, the dryer, the CREC units and local weather data at the datacenter site. The sensor placement of data collection is depicted in Fig. 2. In addition to monitored data, the moisture content of the material, initial moisture content as wet and after drying, was measured sample based manually in intervals during the test campaign. However, the dataset was limited, comprising only 15 pairs of values before and after drying.

## 4. METHODS

Several different machine learning methods have been used in this research to find insights from the data. In addition, to interpret these models, methods of XAI are used.

### 4.1 Machine learning models

In this research, first, a linear regression model based on ordinary least square regression was used in biomass moisture prediction and the model's ability to generalize new

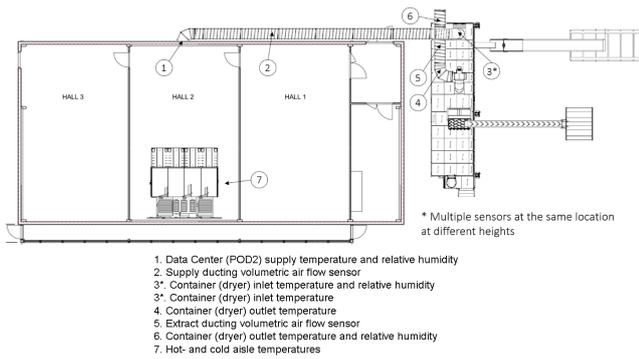


Fig. 2. The data collection points and placement of the sensors at the test site.

data have been inspected based on leave-one-out-cross-validation (LOOCV) (Hastie et al., 2001). Subsequently, five predictive ML methods - LM, GBM, RF, XGBoost and MLP - were employed to predict the absolute humidity of exhaust air, utilizing 10-fold cross-validation during training. For each model except LM, hyperparameter tuning was conducted using a grid-based approach. The tuning process aimed to optimize the models' performance metrics, such as root mean squared error (RMSE), using a cross-validation.

All of these methods are types of supervised machine learning, where the algorithms are trained on labeled data to predict the desired variable. These methods were employed with the goal of identifying the factors that affect biomass end moisture and predicting accurately the exhaust air humidity after one and half hour, which can aid in optimizing the initial settings for the drying process.

GBM, RF, and XGBoost are tree-based ensemble methods (Zhou, 2014). They work by combining the predictions of multiple individual models (trees in the case of RF and GBM, and boosted trees in the case of XGBoost) to make a final prediction. In RF, each tree is built independently using a random subset of features and samples, and the final prediction is made by averaging or voting over the predictions of all trees (Breiman, 2001). In GBM and XGBoost, trees are built sequentially, with each new tree trained to correct the errors made by the previous trees. (Hastie et al., 2001; Chen and Guestrin, 2016) MLP is a type of artificial neural network (ANN) that consists of multiple layers of interconnected neurons, including an input layer, one or more hidden layers, and an output layer. Each neuron in an MLP is connected to every neuron in the adjacent layers, allowing for complex non-linear mappings between input and output features (Bishop, 1995).

#### 4.2 Visualization with XAI

Explainable artificial intelligence (XAI) methods have been used to enhance the understandability of the modeling results (Goebel et al., 2018). With XAI methods, the interpretation of the black box models can be increased (Apley and Zhu, 2019). MLPs are generally considered to be less interpretable compared to tree-based models like RF, GBM, and XGBoost. This is because the internal workings of a neural network are highly complex and opaque, making it difficult to understand how individual features contribute to the model's predictions. In contrast,

decision trees are more transparent and can provide insights into which features are important for making predictions.

Transparency can be increased by providing information about the strength of the importance of each feature of the model regardless of the modeling method used. In this study, model-independent SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) and Accumulated Local Effects Plots (ALE) (Friedman, 2002) techniques are used. SHAP based on the concept of Shapley values from cooperative game theory to provide a unified approach for explaining the predictions of any machine learning model. They quantify the contribution of each feature by systematically varying its value while keeping others constant and comparing the model's prediction to a baseline. With ALE, the average effect of features on the predictions of ML model can be visualized. The interactions between features are also important, and the strength of interactions can be estimated.

## 5. RESULTS

In this research, there are two study cases, each with its own dataset. The first study case focuses on predicting biomass moisture, while the second study case focuses on predicting the humidity of the exhaust air.

### 5.1 Dataset

In these two cases, about two weeks' dataset were available, although with some limitations. In the biomass moisture study case, the dataset comprises 15 pairs of biomass moisture measurements, recorded manually as spot checks. The pairs were selected with a minimum one-hour delay between start and end moisture readings during the period of time.

In the absolute humidity of exhaust air study case, the data is divided iteratively into training and test sets using 10-fold cross-validation. The independent features used for model training, are 5-minute average values and there are 12 features each containing 140 values. Features are listed in Table 1. The dependent variable, exhaust air absolute humidity, represents the absolute humidity after one and a half hours, relative to the starting state, which corresponds to the drying time of one portion of the biomass.

### 5.2 Study case: Biomass moisture

Rather than predicting end moisture directly, a linear regression model was employed to predict the difference between start and end moisture. Variable selection during modeling resulted in a final model with 9 variables predicting this moisture difference. Figure 3 displays the actual versus predicted moisture differences, while Figure 4 provides detailed model results. Despite an 89% explanation of variance ( $R^2$ ), the adjusted  $R^2$  (0.69) suggests the presence of non-significant variables, notably POD2 supply temperature ( $p = 0.01$ ), supply flow ( $p = 0.01$ ), and outdoor absolute humidity ( $p = 0.02$ ), identified as significant based on p-values. The coefficient plot, with 95% confidence intervals, indicates considerable uncertainty in variables, with some intersecting the reference line at 0,

Table 1. Description of the independent variables

Name	Description
container_supply_hum	incoming air humidity measured in the end of air pipe (%)
container_supply_temp	incoming air temperature measured in the end of air pipe (°C)
dryer_fan_speed	incoming air flow rate to dryer(%)
dryer_feed_rate	material feeding rate to the dryer (%)
humidity_outdoor_absolute	the absolute value of outdoor air humidity ( $g/m^3$ )
humidity_outdoor_relative	the relative value of outdoor air humidity (%)
POD2_supply_hum	incoming air humidity from the data center to air pipe (%)
POD2_supply_temp	incoming air temperature from the data center to air pipe (°C)
precipitation_quantity_absolute	the absolute quantity of half on hour cumulative precipitation (mm)
Supply_air_temp	incoming air temperature from pipe to dryer (°C)
Supply_air_relative_humidity	incoming relative air humidity from pipe to dryer (%)
supply_flow	Air flow from data center to dryer measured in the middle of the pipe ( $m^3/s$ )

signifying insignificance in Fig. 5. Notably, absolute precipitation quantity and relative exhaust air humidity were non-significant, while others influenced biomass moisture differences. The model performance was evaluated using LOOCV, yielding an  $R^2$  value of only 0.57. The generalization of this model appears to be quite poor.

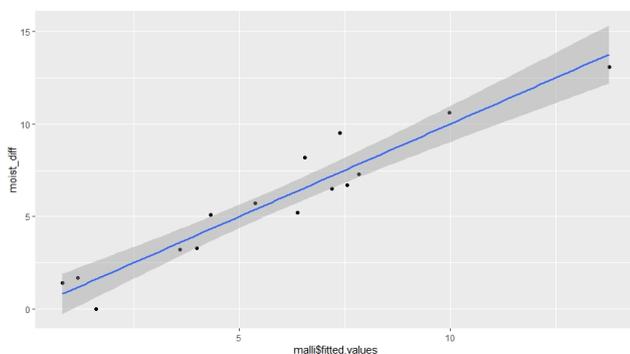


Fig. 3. A comparison between observed and predicted biomass moisture differences with LM.

**MODEL INFO:**  
 Observations: 15  
 Dependent Variable: moist\_diff  
 Type: OLS linear regression

**MODEL FIT:**  
 $F(9, 5) = 4.51$ ,  $p = 0.06$   
 $R^2 = 0.89$   
 Adj.  $R^2 = 0.69$

Standard errors: OLS

	Est.	S. E.	t val.	p
(Intercept)	-26.86	7.22	-3.72	0.01
POD2_supply_temp	31.20	7.53	4.15	0.01
container_supply_hum	44.29	15.62	2.84	0.04
container_exhaust_hum	-10.72	5.28	-2.03	0.10
supply_flow	22.37	5.38	4.15	0.01
supply_air_relative_humidity	22.37	10.53	2.12	0.09
exhaust_air_relative_hum	6.89	5.33	1.29	0.25
exhaust_air_temp	36.89	14.01	2.63	0.05
humidity_outdoor_absolute	-80.51	22.71	-3.55	0.02
precipitation_quantity_absolute	4.59	3.04	1.51	0.19

Fig. 4. LM performance evaluation.

### 5.3 Study case: Absolute humidity of exhaust air

In this study, the absolute humidity of exhaust air is predicted using five ML models: LM, GBM, RF, XGBoost, and MLP. Each model was trained and evaluated using 10-fold cross-validation, and all models except LM underwent hyperparameter tuning via grid search. The LM achieved an average RMSE of 0.468 and average  $R^2$  of 0.799. GBM

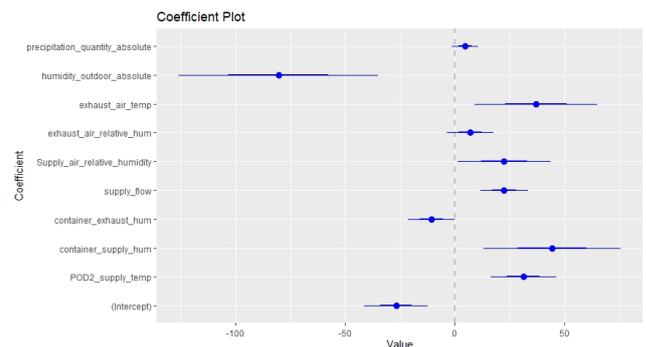


Fig. 5. A coefficient plot of LM with 95% confidence intervals showing the significance of independent features.

Table 2. Comparison of modeling results. Higher  $R^2$  values and smaller RMSE values indicate higher predictive accuracy

Model	$R^2$ Mean $\uparrow$ (sd $\downarrow$ )	RMSE Mean $\downarrow$ (sd $\downarrow$ )
LM	0.799 (0.079)	0.468 (0.113)
GBM	0.880 (0.070)	0.359 (0.107)
RF	0.894 (0.074)	0.345 (0.110)
XGBoost	0.894 (0.075)	0.345 (0.101)
MLP	0.726 (0.005)	0.738 (0.003)

showed the best results with a shrinkage of 0.1, interaction depth of 6, and 200 trees, yielding an average RMSE of 0.359 and an average  $R^2$  of 0.880. RF achieved optimal performance with the number of variables to randomly sample as candidates at each split (mtry) of 4, resulting in an average RMSE of 0.345 and an average  $R^2$  of 0.894. XGBoost performed optimally with a learning rate of 0.3, maximum depth of 3, and 200 rounds, achieving an average RMSE of 0.345 and an average  $R^2$  of 0.894. MLP demonstrated its best performance with a single hidden layer size of 10 and weight decay of 0.01, achieving an average RMSE of 0.738 and an average  $R^2$  of 0.726. All the modeling results, including mean and standard deviation values of  $R^2$  and RMSE for each of five models, are shown in Table 2. As can be seen, the most accurate models are RF and XGBoost based on the RMSE and  $R^2$  values.

The SHAP feature importance plots for both models are presented in Fig. 6 and Fig. 7. In the RF plot, the X-axis values indicate the average absolute SHAP value across all samples, representing the average impact of each feature has on the model's output across different data points. A higher absolute SHAP value suggests that the feature has a stronger influence on the model's prediction. Conversely,

In XGBoost, the X-axis labeled "Gain" denotes the gain in model performance achieved by splitting on each feature during the training process. This gain is calculated based on various metrics such as information gain or reduction in impurity. Features with higher gain values indicate that splitting on those features leads to greater improvements in the model's predictive accuracy. In both plots, the top four features are the same but in a different order. The most important features are the amount of humidity expelled from the data center (POD2\_supply\_hum), absolute humidity of outdoor (humidity\_outdoor\_absolute), absolute quantity of precipitation (precipitation\_quantity\_absolute), and the speed of the dryer fan (dryer\_fan\_speed). In other words, these features exhibit the most significant influence on the target variable, the exhaust air humidity from the dryer.

The correlation between features has been calculated and is shown as a heatmap in Fig. 8. It can be seen that the most correlated feature pairs are those measuring the supply air humidity and temperature. Additionally, the speed of the fan correlates quite strongly (0.80) with the amount of humidity expelled from the data center (POD2\_supply\_hum), the relative outdoor humidity (0.89), and the absolute precipitation quantity (0.87). The RF model will be interpreted in more detail next.

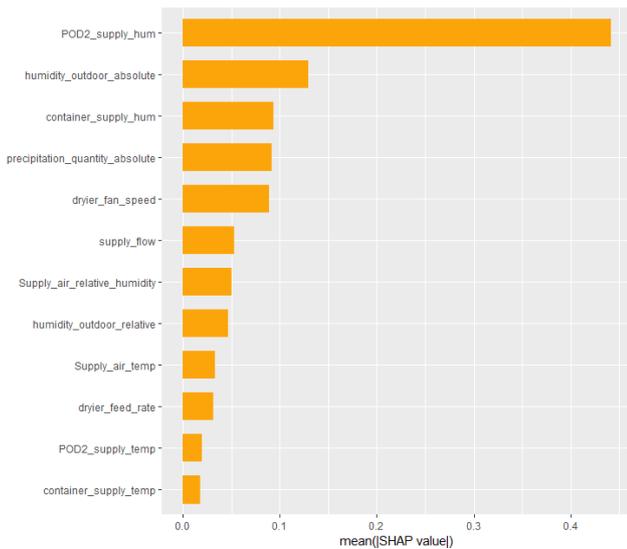


Fig. 6. The SHAP feature importance for RF.

The effect of each feature on the dependent variable is visualized with Accumulated Local Effects (ALE) plots. Figure 9 presents the ALE plot illustrating the influence of the input feature 'POD2\_supply\_hum' on the exhaust air absolute humidity. ALE shows the main effect of the feature at a certain value compared to the average prediction of the exhaust air absolute humidity of RF and also the distribution of data points. As can be seen, the predicted exhaust air absolute humidity is higher when the POD2 supply humidity is below 12.8%. Similarly, a clear boundary can be observed where the feature begins to have a negative effect on the exhaust air absolute humidity in Figs. 10 to 13. These plots show that the exhaust air absolute humidity is higher when the absolute humidity of outdoor (humidity\_outdoor\_absolute) is below  $6.7g/m^3$ , the container supply humidity is below

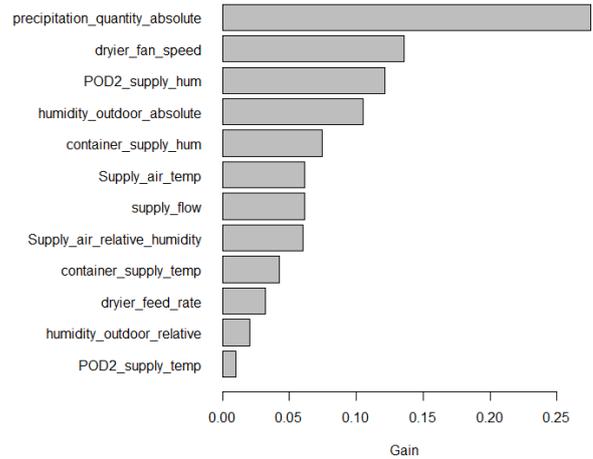


Fig. 7. The SHAP feature importance for XGBoost.

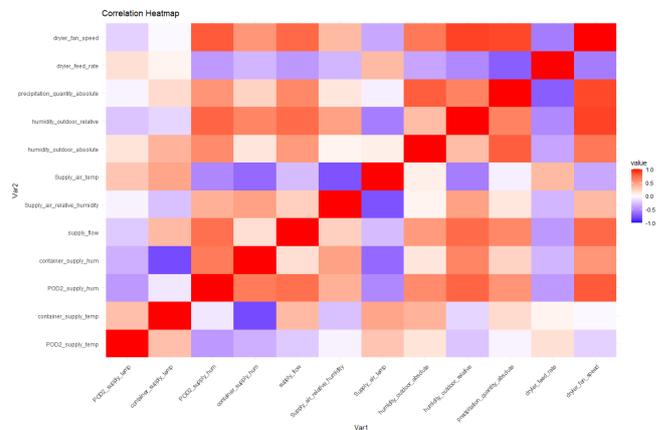


Fig. 8. Correlation heatmap of the independent features.

15.3%, the absolute quantity of precipitation is below 139mm (precipitation\_quantity\_absolute), and the speed of the dryer fan (dryer\_fan\_speed) should be 50% instead of 100%. The modeling results are influenced not only

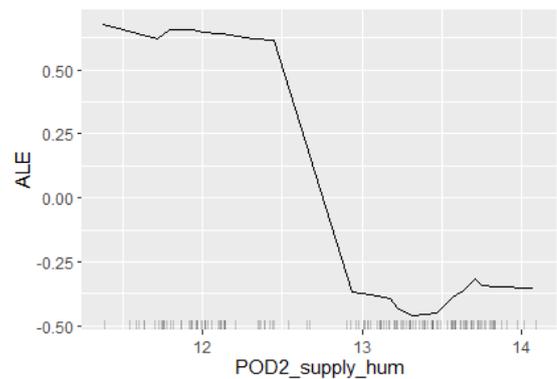


Fig. 9. ALE of the POD2 supply humidity (%) on the predicted humidity of RF (black solid line) are shown on the Y-axis, with the distribution of data points represented by black bars along the X-axis.

by individual features but also by their interactions. In addition to analyzing features independently, it's essential to study their interactions. These interactions are depicted

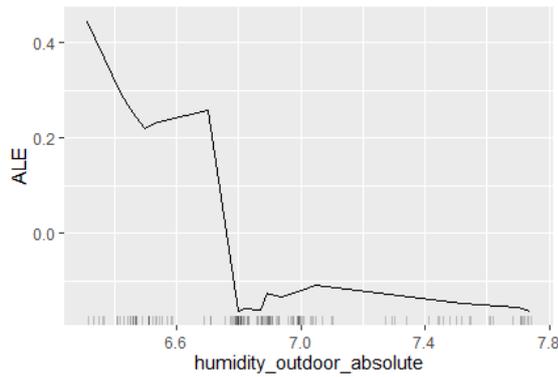


Fig. 10. ALE of the absolute humidity of outdoor ( $g/m^3$ ) on the predicted humidity of RF.

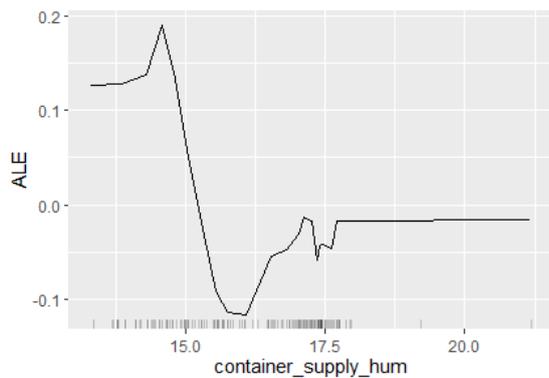


Fig. 11. ALE of the container supply humidity (%) on the predicted humidity of RF.

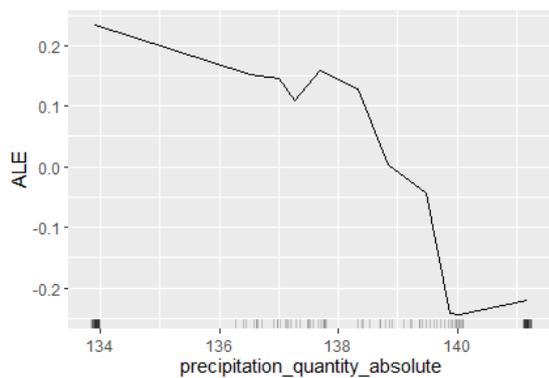


Fig. 12. ALE of the absolute quantity of precipitation (mm) on the predicted humidity of RF.

in Fig. 14, illustrating the strength of interplay between features. Each feature is assigned an interaction strength value, ranging from 0 to 1, indicating the proportion of explained variance of  $f(x)$ . A value of zero denotes no interaction, while one indicates complete dependence on the interaction of the given feature. Notably, POD2 supply humidity, container supply humidity, relative and absolute humidity of outdoor, and absolute precipitation quantity exhibit the highest interaction strengths. Moreover, Figure 15 visually identifies the strongest interaction partners for each feature. For instance, the POD2 supply humidity's primary interaction partner is the absolute humidity of outdoor. Furthermore, interactions between absolute pre-

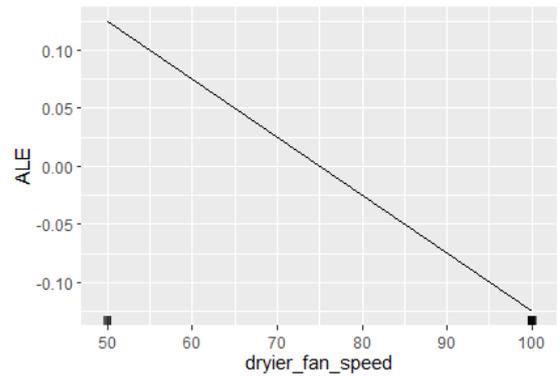


Fig. 13. ALE of the speed of the fan (%) on the predicted humidity of RF.

cipitation quantity, relative humidity of outdoor, container supply air humidity, and the speed of fan are also observed. But overall, the interaction rates are quite low, indicating a relatively low interaction between the features.

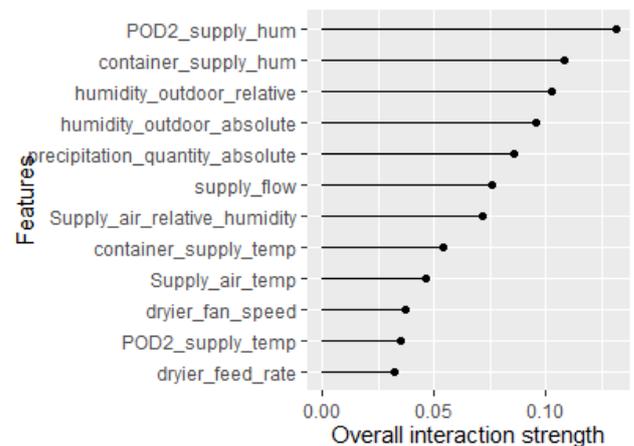


Fig. 14. Overall interaction strengths for each feature independently analyzed, derived from the RF model.

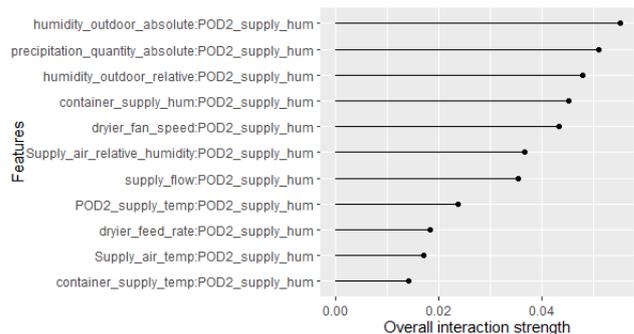


Fig. 15. The 2-way interaction strengths between the most important feature, POD2 supply humidity, and the other features, as determined by the RF model.

For individual predictions, SHAP summary plots provide a visualization of the SHAP values for each feature. A summary of SHAP values illustrates the impact of each feature on the predicted exhaust air absolute humidity in Fig. 16. A positive SHAP value for a feature indicates

that its presence increases the prediction, while a negative SHAP value indicates that its presence decreases the prediction. The length of the bar shows the magnitude of the impact. Longer bars mean a greater impact on the prediction. It's evident that lower POD2 supply humidity, absolute humidity outdoors, and container supply humidity values correspond to higher predicted exhaust air absolute humidity. Conversely, higher values result in lower predicted humidity. As can be seen, with the low POD2 supply humidity, absolute humidity of outdoors and container supply humidity values, the predicted exhaust air absolute humidity is higher. Similarly, with higher values, the predicted humidity is lower.

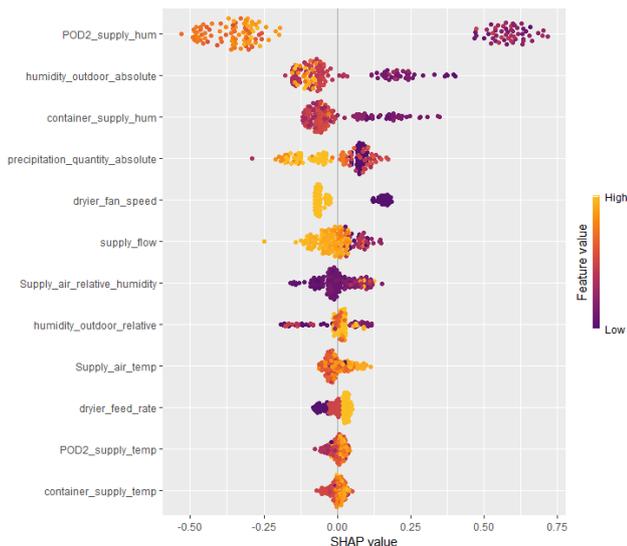


Fig. 16. Summary of SHAP values illustrating the impact of each feature on RF model predictions.

## 6. DISCUSSION

The prediction of biomass end moisture poses a significant challenge due to the inherent difficulty indirectly measuring this target variable. In this study, a small set of initial measurements was measured both before and after the drying process. However, the small size of the dataset and the quality issues associated with the measurements contribute to the overall low modeling accuracy. Consequently, an alternative approach focusing on the exhaust absolute humidity was adopted, as it proves to be more easily measurable. To mitigate the challenges posed by the limited datasets, cross-validation techniques were utilized to enhance the robustness of model evaluation and minimize the potential for overfitting.

The relationship between exhaust air humidity and moisture captured from the material reveals valuable insights into the drying process. Specifically, in the absolute humidity of exhaust air study case, it was assumed that higher humidity levels in the exhaust air corresponded to greater moisture extraction from the biomass, resulting in a drier biomass output. Root cause analysis of exhaust air humidity reveals underlying root causes affecting biomass moisture, as well.

The most important factor behind the humidity expelled from the dryer is the amount of humidity expelled from the

data center (referred to as the POD2 supply humidity), as indicated by the RF feature importance analysis. Additionally, the absolute quantity of precipitation (referred to as precipitation quantity absolute) emerges as another crucial factor, as identified by the XGBoost analysis. Both features rank in the top four in both analyses. Variations in air humidity along the path from the data center to the dryer, as well as the initial supply humidity (referred to as the container supply humidity) at the onset of drying, play significant roles in both feature analyses. In both analyses, the top two most important features include the absolute outdoor humidity and the absolute quantity of precipitation, which are quite strongly correlated based on the heatmap. Therefore, weather conditions, particularly rainy weather, are found to reduce the effectiveness of the biomass drying process due to alterations in the quality of leaked air.

The SHAP summary plots provided valuable insights into the impact of each feature on predicted exhaust air absolute humidity. The analysis revealed that lower POD2 supply humidity, absolute humidity outdoors, and container supply humidity were associated with higher predicted exhaust air absolute humidity, as indicated by positive SHAP values. Conversely, higher values of these features were linked to lower predicted humidity, as evidenced by negative SHAP values. These findings highlight the significance of controlling and optimizing factors such as POD2 supply humidity and outdoor humidity to manage and regulate exhaust air absolute humidity levels during the drying process.

Through data analysis, three threshold values were identified to increase the exhaust absolute humidity: POD2 supply humidity below 12.8%, container supply humidity below 15.3%, and the speed of the dryer fan set at 50% instead of 100%. Additionally, favorable weather conditions related to air humidity were identified: the absolute humidity outdoors should be below  $6.7 \text{ g/m}^3$ , and the absolute quantity of precipitation should be below 139 mm. If these thresholds are exceeded, process settings can be adjusted to optimize biomass drying, resulting in a drier biomass output.

Moreover, weather conditions interact with other independent features, highlighting the importance of improving system air tightness to enhance drying efficiency. Shorter and more airtight pipes between the data center and dryer, along with airtight design of air entry and output for the dryer, can mitigate the adverse effects of humid weather conditions, further optimizing biomass drying effectiveness.

The developed model accurately predicts exhaust air absolute humidity for one hour and a half into future with present settings. Additionally, both the RF and XGBoost models demonstrate strong generalization capabilities, as evidenced by the metrics calculated on an independent test dataset obtained through 10-fold cross-validation. These models explain 89% of the variation in exhaust air absolute humidity using the selected independent features. Nonetheless, further research is needed to refine the model's ability to indirectly predict biomass end moisture. With access to more data for model training, testing, and validation, dynamic modeling could enable long-term

predictions of end moisture and adaptive system control, further optimizing the entire biomass drying process.

## 7. CONCLUSION

It has been shown that the developed models, even with a small dataset, showcased the potential of ML and its capabilities in developing process control for biomass drying, particularly through tree-based ML methods. As the test campaign showed the ambient conditions effect strongly to the dryer's efficiency especially when low temperature drying is done. And when operating a multivariant system like drying in a changing environment there needs to be possibilities to measure and react to the changes. The machine learning could provide a tool for handling these changes by predicting the upcoming changes based on the history data or forecast data e.g. of weather conditions or material initial moisture content changes. More exact evaluation of process parameters variations enables more accurate prediction of e.g., production capacities.

## ACKNOWLEDGEMENTS

The work for this paper was done as part of ArctiqDC project, with financial support from the European Regional Development Fund via the Interreg Nord program. RISE Research Institutes of Sweden is acknowledged for their contribution to data collection.

## REFERENCES

- Apley, D.W. and Zhu, J. (2019). Visualizing the effects of predictor variables in black box supervised learning models.
- Ascher, S., Watson, I., and You, S. (2022). Machine learning methods for modelling the gasification and pyrolysis of biomass and waste. *Renewable and Sustainable Energy Reviews*, 155, 111902.
- Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford University Press.
- Breiman, L. (2001). Random forests. *Mach. Learn.*, 45(1), 5–32. doi:10.1023/A:1010933404324.
- Chai, H., Chen, X., Cai, Y., and Zhao, J. (2019). Artificial neural network modeling for predicting wood moisture content in high frequency vacuum drying process. *Forests*, 10(1). doi:10.3390/f10010016.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, 785–794. Association for Computing Machinery, New York, NY, USA. doi:10.1145/2939672.2939785.
- Friedman, J.H. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal.*, 38(4), 367–378. doi:10.1016/S0167-9473(01)00065-2.
- Gebreegziabher, T., Oyedun, A.O., and Hui, C.W. (2013). Optimum biomass drying for combustion – a modeling approach. *Energy*, 53, 67–73. doi:10.1016/j.energy.2013.03.004.
- Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., and Holzinger, A. (2018). Explainable ai: The new 42? In A. Holzinger, P. Kieseberg, A.M. Tjoa, and E. Weippl (eds.), *Machine Learning and Knowledge Extraction*, 295–303. Springer International Publishing, Cham.
- Hagras, H. (2018). Toward human-understandable, explainable ai. *Computer*, 51(9), 28–36. doi:10.1109/MC.2018.3620965.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York.
- He, S., Wang, G.A., and Li, L. (2009). Quality improvement using data mining in manufacturing processes. In J. Ponce and A. Karahoca (eds.), *Data Mining and Knowledge Discovery in Real Life applications*, 357–372. I-Tech.
- James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Li, G., Li, Z., Yin, T., Ren, J., Wang, Y., Jiao, Y., and He, C. (2022). Drying biomass using waste heat from biomass ash by means of heat carrier. *BioResources*, 17(3), 5243–5254.
- Li, H., Chen, Q., Zhang, X., Finney, K.N., Sharifi, V.N., and Swithenbank, J. (2012). Evaluation of a biomass drying process using waste heat from process industries: A case study. *Applied Thermal Engineering*, 35, 71–80. doi:10.1016/j.applthermaleng.2011.10.009.
- Lundberg, S.M. and Lee, S.I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, 30, 4765–4774. Curran Associates, Inc.
- Onsree, T. and Tippayawong, N. (2021). Machine learning application to predict yields of solid products from biomass torrefaction. *Renewable Energy*, 167(C), 425–432.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.
- Wahlroos, M., Pärssinen, M., Rinne, S., Syri, S., and Manner, J. (2018). Future views on waste heat utilization – case of data centers in northern europe. *Renewable and Sustainable Energy Reviews*, 82, 1749–1764. doi:10.1016/j.rser.2017.10.058.
- Zhou, Z.H. (2014). Ensemble methods. *Combining pattern classifiers*. Wiley, Hoboken, 186–229.