

Emission prediction from waste mask pellet combustion: evaluating multiple linear regression and advanced machine learning approaches

Alexander BACKA^{1*}, Nikola ČAJOVÁ KANTOVÁ¹, Róbert CIBULA¹ and Alexander ČAJA²

Authors' affiliations and addresses:

¹ Research Centre, University of Žilina,
Univerzitná 8215/1, 010 26 Žilina, Slovakia
e-mail: alexander.backa@uniza.sk
e-mail: nikola.cajovakantova@uniza.sk
e-mail: robert.cibula@uniza.sk

² Department of Power Engineering, Faculty of
Mechanical Engineering, University of Žilina,
Univerzitná 8215/1, 010 26 Žilina, Slovakia
e-mail: alexander.caja@fstroj.uniza.sk

*Correspondence:

Alexander Backa, Research Centre, University of
Žilina, Univerzitná 8215/1, 010 26 Žilina,
Slovakia
tel.: +421 41 513 7610
e-mail: alexander.backa@uniza.sk

Acknowledgement:

This publication has been produced with the
support of VEGA No. 1/0150/22: Energy
utilization of produced waste in connection with
the COVID-19 pandemic through pellets as an
alternative fuel, VEGA No. 1/0311/26:
Autonomous control of the combustion process in
small heat sources using artificial
intelligence and KEGA No. 026ŽU-4/2025:
Innovative approaches to teaching
thermomechanics and fluid mechanics.

How to cite this article:

Backa, A., Čajová Kantová N., Cibula, R. and
Čaja A. (2026) Emission prediction in the energy
utilization of pellets from waste masks: a machine
learning analysis. *Acta Montanistica Slovaca*,
Volume 31 (2), 438-448

DOI:

<https://doi.org/10.46544/AMS.v31i2.09>

Abstract

Growing concern over sustainable waste management and emissions control has prompted increased interest in alternative fuels derived from non-traditional waste materials. This study compares the development of predictive multiple linear regression (MLR) models with advanced machine learning (ML) models to estimate CO, NO_x, and particulate matter (PM) emissions during the combustion of alternative pellet fuels. The experimental fuel blends combined disposable masks and FFP2 respirators with spruce sawdust at varying concentrations (0%, 2%, 3%, 10%, 20%, and 50%). A training dataset was created using data from all blends except the 10% concentration. Emission predictions were performed using both multiple linear regression and advanced machine learning approaches, incorporating key input variables: the proportion of masks and respirators in the fuel, supplied airflow, ambient temperature, chimney draft, chimney temperature, O₂ content in flue gases, and heat output. The 10% mask-content fuel blend dataset was reserved for validation and unseen during model development. Advanced machine learning models demonstrated superior predictive accuracy in this validation test, achieving reductions of 31%, 50.51%, and 94.05% in root mean square error for CO, NO_x, and PM, respectively, compared to multiple linear regression. Despite lower precision, multiple linear regression yielded closed-form linear equations for simple emission predictions across fuel compositions. The findings suggest that machine learning techniques can enhance emission prediction accuracy. These models enable emission forecasting for various fuel mixtures, reducing the need for individual measurements.

Keywords

prediction models, combustion, small heat source, alternative fuels, regression



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Heat sources are integral to nearly every household, workplace, and residential complex. These systems often burn various fuels, with the most common used for heating being coal, natural gas, oil, renewable sources (such as biofuels), and waste (Martinopoulos, Papakostas & Papadopoulos, 2018). Beyond direct combustion routes, several pathways for valorizing waste streams are being explored, including material recovery and energy recovery strategies in the built environment (Latosinska et al., 2023). However, their combustion generates emissions that negatively impact the environment and human health. This necessitates advancements in traditional combustion technologies and the development of innovative approaches to reduce emissions while improving combustion efficiency (Nussbaumer, 2003; Permchart & Kouprianov, 2004).

With approximately 4.2 million units in operation, pellet stoves represent Europe's most widely used pellet heating system, surpassing boilers, which are estimated at around 1.2 million units (Gauthier & Avagianos, 2021). While emissions from pellet combustion are generally lower than those from burning wood logs, they still contribute to air pollution (Ozgen et al., 2014).

There are many ways of adjusting the production of emissions from combustion. One of the options is to enhance the combustion process during operation by monitoring key combustion parameters and adjusting the combustion device accordingly based on the monitoring results (Xia, Zhang, Tang, & Pan, 2023; Zha et al., 2023).

For such a purpose, measuring quantities is not always necessary, but predicting them is also possible. When predicting combustion parameters, in most existing mathematical models, the approach was to solve (differential) equations describing the course of combustion. Most existing mathematical models for predicting combustion parameters rely on solving differential equations that describe the combustion process (Boriouchkine & Jämsä-Jounela, 2016; Boriouchkine, Zakharov, & Jämsä-Jounela, 2012; Porteiro et al., 2006; Van Der Lans et al., 2000). The key differences among these models lie in the methods used to solve these equations and the development of entirely new modeling techniques tailored to specific research objectives (Zheng et al., 2020). However, many of these models are computationally complex and require significant transformations when applied to different conditions (Morari & Lee, 1999).

A common approach to solving combustion-related differential equations is Computational Fluid Dynamics (CFD), a numerical method for simulating fluid and gas flow. CFD enables the modeling of complex physical processes, including combustion, by simulating airflow, fuel distribution, and gas concentrations within burners and exhaust channels. This allows researchers and engineers to test various scenarios and optimize combustion system designs to improve efficiency and reduce emissions (Zadravec et al., 2020; Adamczyk et al., 2014). However, CFD simulations can be computationally demanding and time-intensive, particularly for multiphase combustion systems. High computational costs and specialized hardware/software requirements may become significant limitations in such cases. As a result, there is growing interest in developing faster and more efficient modeling approaches or hybrid techniques that combine different methods (Panchigar et al., 2022).

Although various kinetic and equilibrium models exist for describing combustion processes, they are often unsuitable for real-time application in operational combustion systems. A promising alternative is the use of artificial intelligence (AI). Machine learning (ML), a subfield of AI, employs algorithms that learn from diverse data sources, including numerical values, images, text, and audio. Regression models are widely used for quantitative prediction tasks, including forecasting real estate prices, estimating costs, predicting energy consumption, and health assessments. However, while regression models can provide valuable insights, it is crucial to recognize their limitations and interpret results cautiously. ML-based regression models can analyze variable interrelationships and be used for predictive purposes.

Machine-learning approaches are increasingly used across diverse energy sectors, including the evaluation of indoor environmental quality and thermal comfort in buildings (Belany et al., 2024; Majewski et al., 2020; Orman et al., 2022). In the context of combustion, machine learning regression models can predict parameters such as temperature, emissions, particulate matter (PM) generation, and heat source power (Cao, Xin & Yuan, 2016; Čajová Kantová, 2024a; Ihme, Chung & Mishra, 2022; Mutlu & Yucel, 2018). For instance, Böhler et al. (2021) investigated an emission-limited model predictive control (MPC) system for a small-scale biomass furnace, aiming to minimize carbon monoxide (CO) and nitrogen oxides (NO_x) emissions while maintaining combustion efficiency. Their study demonstrated that predictive control strategies significantly reduced CO emissions, particularly under partial-load conditions, achieving up to a 4-fold reduction compared to conventional control methods. This model-based approach successfully integrated an emission estimation model, optimizing combustion parameters for robust, fuel-flexible control and improved environmental performance without extra sensors.

This study investigates the use of multiple linear regression (MLR) and advanced machine learning (ML) techniques for predicting CO, NO_x, and PM emissions during the combustion of alternative pellets. Accurate emission prediction is crucial when fuel or combustion conditions vary, enabling emission estimation without direct measurement. The pellets used in this study were produced by mixing spruce sawdust with crushed disposable face masks and FFP2 respirators at varying concentrations (0%, 2%, 3%, 10%, 20%, and 50%). Data from the 10% mask-content sample were reserved for validation, serving as an independent test on the same heat

source. The goal was to develop predictive models that estimate emissions for different alternative fuel compositions, reducing the need for extensive direct measurements. Related experimental characterization of mask-based spruce pellets in a small-scale boiler has been reported previously (Čajová Kantová et al., 2024c).

ML models not only enable emission prediction but also have the potential to control combustion actuators and optimize heat-source operation, thereby improving efficiency. While experimental measurements were necessary for model development and validation in this study, future research could refine these models to require fewer measurements or inputs. Machine learning integration enhances emission-prediction accuracy and enables real-time analysis, both of which are valuable for dynamic combustion systems.

Material and Methods

Disposable masks and FFP2 respirators were crushed and added in equal proportions at concentrations of 0%, 2%, 3%, 10%, 20%, and 50% to spruce sawdust, along with a control sample containing no FFP2 or disposable masks (Fig. 1). The moisture content of the produced pellets was determined prior to combustion. Because the spruce sawdust provided the dominant hygroscopic fraction, the resulting moisture levels varied only slightly across the fuel blends. The measured moisture contents were 5.18% (100/0), 5.44% (98/2), 6.84% (97/3), 7.62% (90/10), 5.29% (80/20), and 5.28% (50/50). The masks and FFP2 respirator fragments exhibited moisture contents below 1%. Therefore, their contribution to the overall pellet moisture was negligible. All produced blends were thus within the typical moisture range recommended for stable combustion in small-scale pellet boilers.



Fig. 1. Alternative fuel pellets with crushed disposable face masks and FFP2 respirators at 0%, 2%, 3%, 10%, 20%, and 50% content

This mixture was used to produce 6 mm diameter pellets. These pellet samples were combusted in a LOKCA ÚSPOR 18 AUTOMAT automatic pellet boiler (Fig. 2). The boiler has a maximum rated heat output of 18 kW and is equipped with a retort burner.



Fig. 2. Small pellet boiler with retort burner.

Fuel dosing occurred at a constant rate of approximately 4.8 kg/h, pulsed with a 15-second feed interval and a 25-second standby interval. Airflow was maintained at a constant rate with minimal observed fluctuations. The generated heat was transferred via a heat exchanger to a cooling tower. A simplified schematic of the main measured variables for prediction purposes is shown in Fig. 3. For each fuel sample, 2 measurements were performed lasting 30 minutes after the heat output and emission values had stabilized. Data collection was performed every 10 seconds.

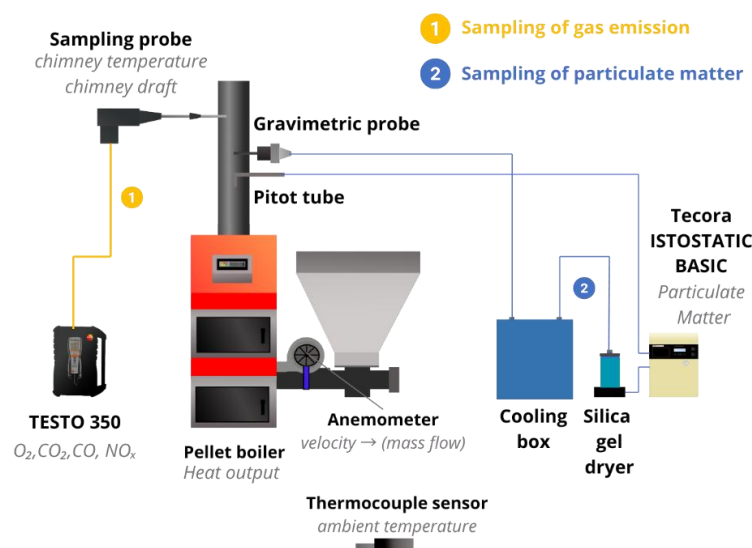


Fig. 3. The schematic diagram of parameter measurement on the heat source.

The air mass flow rate was calculated by measuring air velocity at the fan inlet with an FVAD 15 MA1 anemometer, and then adjusting for air density based on temperature and humidity. The ambient temperature was measured using thermocouples. These measurements were taken in the environment surrounding the combustion process. The chimney draft was continuously monitored using a differential pressure sensor with a maximum range of 100 Pa and an accuracy of 0.5%. This device measured the pressure difference to determine the chimney's draft during combustion. The chimney temperature was measured using a Type K (NiCr-Ni) thermocouple, located at a specific point 1.5 meters above the heat source in the chimney. This sensor provided accurate temperature readings for the flue gases. Gas emissions in the flue gases were measured using a gas analyzer, the TESTO 350, which detects gas concentrations. The boiler's heat output was determined by measuring the thermal difference between the inlet and outlet temperatures of the heating medium. This calculation relied on the calorimetric formula, which includes the mass flow rate of water and its specific heat capacity. Gravimetric measurement of particulate matter used a complex sampler (according to ISO 9096:2017). The gravimetric probe was connected to a Tecora ISOSTACK BASIC, a Pitot tube, a cooling box, and a dryer.

This study developed simple, easily interpretable multiple linear regression (MLR) models, enabling the creation of equations with coefficient estimates for individual variables. MLR is a statistical method for modeling the relationship between a dependent variable and multiple independent variables. It extends simple linear regression by incorporating two or more predictors, allowing for a more comprehensive analysis of complex systems (Jobson, 1991). These models were created in Excel using the Data Analysis Regression add-in. The general form of the linear regression equation is:

$$y_n = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (1)$$

Where y_n is the specific target variable, b_0 is the intercept, b_1 - b_n are the regression coefficients, and x_n is the individual input variable, as listed in Table 1.

While MLR provides a straightforward, interpretable approach to modeling relationships between variables, more advanced ML regression techniques offer greater flexibility for capturing complex, nonlinear dependencies. ML regression models are powerful tools for predicting continuous variables by identifying patterns in data. These models range from simple linear regression to more sophisticated techniques capable of handling intricate relationships.

Using MATLAB's Regression Learner app, various regression models were tested on the same datasets, including regression trees, Gaussian process regression models, support vector machines, kernel approximation, ensembles of regression trees, neural network regression, and various linear regression models. These models are widely used in predictive analytics, each offering different trade-offs in terms of interpretability, computational complexity, and accuracy. The input parameters for model development, including the range of measured values, are provided in Table 1. The percentage of added face masks and respirators affects fuel composition, while supplied airflow, chimney draft, and O_2 content regulate combustion efficiency. Ambient and chimney temperatures impact reaction dynamics, and heat output reflects overall combustion performance. The selection of specific quantities was based on knowledge of the influence of the given parameter, measurements made, the availability of measuring devices for the boiler and fuel, in connection with previous research conducted by Kantová et al. (2023, 2024), and on the simplicity of constructing the models. A parameter inherently affecting

combustion is fuel moisture content. Moisture influences ignition delay, flame stability, temperature profiles, burnout quality, and consequently CO formation and PM generation. Previous studies in small-scale systems support the roles of fuel moisture and air management in shaping CO and NO_x trends (Lukáč et al., 2019; Lukáč et al., 2020). Higher moisture content typically lowers combustion temperature due to the latent heat required for water evaporation, potentially increasing CO emissions and altering PM formation pathways. For stable boiler operation, most pellet boiler manufacturers recommend a moisture content of approximately 6–10% for wood-based pellets, ensuring reproducible feeding behaviour, stable flame formation, and predictable emission characteristics (EN ISO 17225-2; typical industry requirements). Because the waste-mask additives were dry polymeric materials, their moisture contribution was negligible. Only unused and uncontaminated masks and FFP2 respirators were used in the experiments. Elemental and inorganic analyses show that these materials contain only trace inorganic elements and very low ash, indicating no detectable formation of toxic chlorinated compounds. In practical waste scenarios, however, discarded masks may contain organic or inorganic residues (cosmetics, dust), which could locally increase volatile release and thus affect CO or PM levels. Such contamination was not present in our samples and therefore, was not included as a predictor in the regression and ML models. If contamination can be quantified in future studies, it may be incorporated as an additional input variable. A more detailed characterisation of these mask/sawdust pellets is provided in a companion study (Čajová Kantová et al., 2024b).

Since only 2 fuel mixtures are considered in the models, for simplicity, instead of fuel composition and calorific value, a variable characterizing the percentage of added face masks and respirators was included. In total, the training dataset comprised nearly 1,800 data points, while the independent validation dataset comprised slightly over 460.

Tab. 1. Descriptive statistics of feature and target variables for training and validation datasets.

Corresponding coefficient	Parameter	Features	Minimum value	Maximum value	Average value + SD
b ₁	x ₁	Percentage of added face masks and respirators (%)	0	50	
b ₂	x ₂	Mass flow rate (kg/h)	99.9	135.6	114.9 ± 5.9
b ₃	x ₃	Ambient temperature (°C)	20.7	22.8	22.2 ± 0.4
b ₄	x ₄	Chimney draft (Pa)	3.8	24.5	12.7 ± 2.2
b ₅	x ₅	Chimney temperature (°C)	147.8	214.5	187.9 ± 11.0
b ₆	x ₆	O ₂ content in flue gases (%)	8.9	16.7	13.5 ± 1.44
b ₇	x ₇	Heat output (kW)	5.8	15.7	11.4 ± 1.99
Target variables					
	y ₁	Carbon monoxide CO (mg/m ³)	236.3	6417.6	1317.8 ± 792.0
	y ₂	Nitrogen oxides NO _x (mg/m ³)	48.6	307.9	109.7 ± 31.1
	y ₃	Particulate matter PM (mg/m ³)	28.3	527.2	193.4 ± 221.9

When comparing prediction models, the root mean square error (RMSE) is often used as an evaluation metric (Baruah et al., 2017; Elmaz et al., 2020). The closer the RMSE value is to zero, the more accurate the proposed model, indicating that the predicted data closely matches the actual input data. RMSE quantifies the average deviation between predicted and actual values. It is easy to interpret and penalizes large errors more heavily than small ones. It is calculated as:

$$RMSE = \sqrt{(\sum(P_i - A_i)^2 / n)} \quad (2)$$

Where:

- P_i is the predicted value for the i -th observation
- A_i is the actual value for the i -th observation
- n is the number of observations

Another important metric is the coefficient of determination (R^2), which quantifies the proportion of the variance in the dependent variable explained by the independent variables in the model. R^2 , ranging from 0 to 1, represents the proportion of dependent variable variance explained by the model. 0 indicates no explained variance, while 1 signifies perfect explanation (Baruah et al., 2017; Elmaz et al., 2020). The formula for calculating R^2 is:

$$R^2 = 1 - (SS_{res} / SS_{tot}) \quad (3)$$

Where:

- SS_{res} is the sum of squares of the residuals (the difference between the actual values and the predicted values)
- SS_{tot} is the total sum of squares (the difference between the actual values and the mean of the actual values)

In addition to RMSE, the mean absolute percentage error (MAPE) is also given, which expresses the average relative deviation between predicted and measured values. MAPE is defined as:

$$MAPE = (100/n)\sum(P_i - A_i)/A_i \quad (4)$$

These evaluation metrics provide a quantitative assessment of model accuracy, but they do not reveal the relative importance of individual input features in shaping predictions. To address this, Shapley importance was applied. Shapley importance, derived from cooperative game theory, is a widely used method for interpreting machine learning models. It assigns an importance value to each feature by considering its marginal contribution across all possible feature combinations. This ensures a fair and consistent distribution of importance scores, making it particularly useful for complex, nonlinear models. In this study, Shapley importance graphs were generated for a basic linear multiple regression model and also for the selected most successful validation model among advanced ML models. These analyses help identify key predictors, enhance model transparency, and support explainable AI applications.

Results and discussion

Basic multiple linear regression models were developed to predict CO, NO_x, and PM emissions from burning pellets composed of spruce sawdust and crushed disposable face masks/FFP2 respirators (0%, 2%, 3%, 20%, and 50% concentrations). The coefficients and statistical metrics derived from these models, following validation on a 10% dataset, are presented in Table 2.

Tab. 2. Statistical characteristics and coefficients of basic multiple linear regression models predicting CO, NO_x and PM emissions.

Calculated coefficient	CO model	NO _x model	PM model
b ₀	10401.666	1262.856	-63.152
b ₁	6.787	-0.436	8.098
b ₂	-20.001	0.668	7.634
b ₃	-445.026	-52.385	50.959
b ₄	30.188	-0.343	-3.005
b ₅	-7.496	-0.431	-9.473
b ₆	323.212	-2.082	-26.526
b ₇	-43.044	3.751	29.336
Model metrics			
R ² - Original training model (without 10% sample)	0.59	0.59	0.87
MAPE - Original training model (without 10% sample)	33.2%	12.9%	104.6%
RMSE - Original training model (without 10% sample)	377.25	21.60	85.77
R ² - Validation test (10% sample)	0.51	0.02	0
MAPE - Validation test (10% sample)	28.2%	28.6%	300.9%
RMSE - Validation test (10% sample)	1017.71	38.53	136.24

The R² values for the CO, NO_x, and PM models in the original training set were 0.59, 0.59, and 0.87, respectively, with the PM model demonstrating a moderate-to-strong fit. The RMSE values for the training models are 377.25 for CO, 21.60 for NO_x, and 85.77 for PM, showing reasonable prediction accuracy for CO and NO_x emissions, with the lowest error seen for NO_x. While the validation test (10% sample) shows a decrease in performance, with R² values of 0.51 for CO, 0.02 for NO_x, and 0 for PM, these results provide valuable insight into the models' variability and suggest that further refinement may be needed to improve their generalization. The RMSE values for the validation test are higher, especially for CO (1017.71), indicating that the models may face challenges in predicting emissions under varied conditions, particularly for CO and PM. For the validation model compared to the trained model, RMSE increased by 169.79% for CO, 78.38% for NO_x, and 58.89% for PM. However, this is because the model is being tested on an independent measurement.

The analysis of the Shapley importance values for the Multiple Linear Regression (MLR) models predicting carbon monoxide (CO), nitrogen oxides (NO_x), and particulate matter (PM) emissions reveals distinct influences of key combustion parameters on each pollutant (Fig. 4).

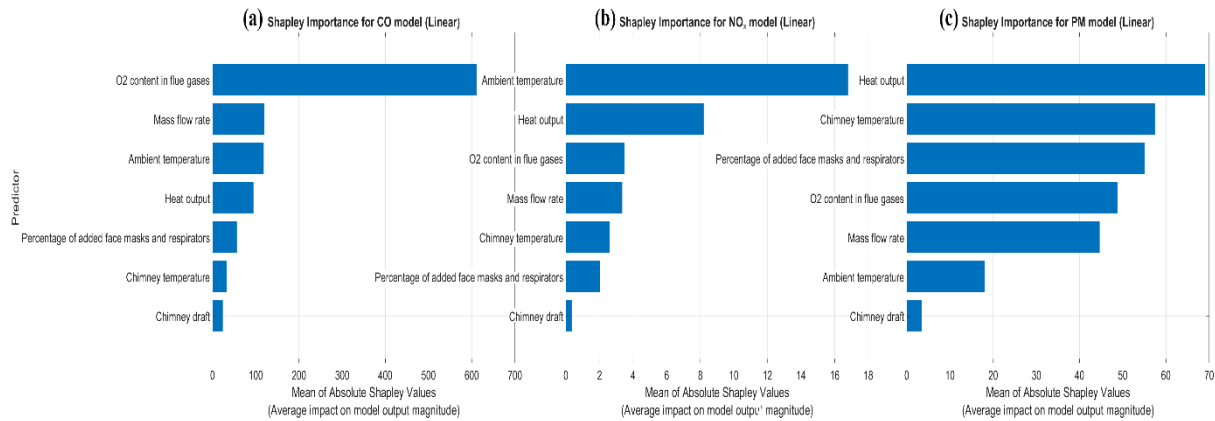


Fig. 4. Shapley importance for CO (a), NO_x (b), and PM (c) prediction using multiple linear regression models.

For the CO model, the O₂ content in flue gases is the dominant factor affecting CO emissions. This is consistent with combustion principles, where an increase in oxygen availability leads to more complete combustion, thereby reducing CO formation. Other influential factors for the MLR model include mass flow rate, ambient temperature, and heat output, albeit with a significantly lower impact than O₂ content. Parameters related to chimney conditions, such as chimney temperature and draft, have little influence, suggesting that post-combustion factors have little effect on CO levels.

In the NO_x model, the most critical predictor is ambient temperature. Heat output and O₂ content in flue gases are key contributors, influencing combustion temperature and oxidation conditions. Mass flow rate and chimney temperature have a moderate impact, while chimney draft and mask/respirator percentage have minimal influence.

The PM model shows a strong correlation between combustion intensity and particulate emissions, with heat output as the most significant factor. Higher heat output suggests increased fuel-burning rates, which can lead to greater particulate formation. Chimney temperature and the percentage of added face masks and respirators also play substantial roles, possibly reflecting variations in combustion efficiency. O₂ content, mass flow rate, and ambient temperature contribute to a lesser extent, while the chimney draft has the lowest impact. The importance of heat output and chimney temperature suggests that both combustion conditions and post-combustion factors influence PM emissions (in the MLR model).

Overall, these findings validate the MLR models' ability to capture key combustion-related processes that affect emissions.

Table 3 presents the statistical characteristics of the obtained advanced machine learning models used to predict CO, NO_x, and PM emissions. The metrics for the most successful test models as well as the most successful validation models are presented here.

Tab. 3. Statistical characteristics of advanced machine learning models predicting CO, NO_x, and PM emissions.

Model type	CO model	NO _x model	PM model
Selected the most successful test models (by lowest RMSE value)	Gaussian Process Regression (Exponential GPR)	Ensemble (Bagged Trees)	Ensemble (Bagged Trees)
R ² - Original training model (without 10% sample)	0.87	0.98	1.00
RMSE - Original training model (without 10% sample)	213.81	5.30	2.32×10 ⁻¹²
MAPE - Original training model (without 10% sample)	15.2%	3.6%	0.0%
R ² - Validation test (10% sample)	0.39	0	0
RMSE - Validation test (10% sample)	756.93	31.76	8.11
MAPE - Validation test (10% sample)	28.6%	23.0%	18.5%
Selected the most successful validation models (by lowest RMSE value)	Support Vector Machine (Cubic SVM)	Efficient Linear (Least Squares)	Tree (Fine Tree)
R ² - Original training model (without 10% sample)	0.82	0	1.00
RMSE - Original training model (without 10% sample)	251.52	33.52	2.42×10 ⁻¹²
MAPE - Original training model (without 10% sample)	17.6%	17.5%	0.0%
R ² - Validation test (10% sample)	0.47	0	0
RMSE - Validation test (10% sample)	702.34	19.07	8.11
MAPE - Validation test (10% sample)	27.5%	13.7%	18.5%

The selected most successful models for predicting CO, NO_x, and PM emission prediction models show substantial improvement over the original training models. Focusing first on the most successful training models,

the Gaussian Process Regression model achieved an R^2 of 0.87 and an RMSE of 213.81 for CO emissions, indicating a strong fit and relatively low prediction error. For NO_x , the Ensemble model performed excellently, with an R^2 of 0.98 and an RMSE of 5.30, demonstrating high accuracy. The PM model, also utilizing the Ensemble method, reached an impressive R^2 of 1.00 and an RMSE close to zero (2.32×10^{-12}), indicating near-perfect prediction accuracy in the training phase. Turning to the most successful validation models, while validation tests showed a decrease in performance compared to training, the models still offered valuable insights. The best validation model for CO was the Support Vector Machine with an R^2 of 0.47 and an RMSE of 702.34. For NO_x , the Efficient Linear model achieved an R^2 of 0 and an RMSE of 19.07, suggesting challenges in validation but still providing some predictive value. Finally, for PM emissions, the Fine Tree model had an R^2 of 0 and an RMSE of 8.11.

The study conducted by Ross-Veitia et al. (2024) evaluated CO and NO_x emissions, along with CO_2 and temperature, from steam boilers using advanced machine learning models and MLR, identifying the Gradient Boosting Regressor as the most accurate predictor, with an RMSE of 2.86 and $R^2 = 0.998$. In contrast, MLR predictions in this study yielded an RMSE of 24.76 and an R^2 of 0.78. Key input variables affecting emissions included ambient temperature, working pressure, steam production, and fuel type. In comparison, our study achieved similarly strong results for the NO_x emissions validation model using MLR, with RMSE = 21.6 and $R^2 = 0.59$. Furthermore, for PM prediction using advanced machine learning models, the RMSE was nearly zero, and the R^2 value reached 1.00, indicating exceptionally high predictive accuracy.

Li et al. (2022) developed a convolutional neural network (CNN)-based model with a channel attention mechanism for predicting NO_x emissions from a coal-fired boiler, achieving high accuracy with RMSE values of 3.03 mg/Nm^3 and 3.24 mg/Nm^3 . The model incorporated 55 input variables, including boiler load, total airflow, fuel flow, steam parameters, oxygen concentration, and various airflow rates, demonstrating strong predictive performance and improved computational efficiency. In contrast, our proposed NO_x model achieves comparable accuracy while using a significantly smaller subset of input parameters. This reduction in input dimensionality not only simplifies the model architecture but also potentially improves its generalization capabilities and reduces the risk of overfitting.

The advanced machine learning models achieved almost perfect predictions of PM emissions during the original training stage, yielding excellent overall results. The RMSE values for the most successful validation models increased by 179.24% for CO and decreased by 43.11% for NO_x (indicating an improvement in accuracy). The RMSE value for PM increased by 335.12 trillion %. The drastic rise in PM RMSE is caused by the initial value being extremely low (almost zero). Even a small absolute increase in RMSE from such a low base translates into a large percentage increase.

The Shapley importance values for the machine learning models predicting carbon monoxide (CO), nitrogen oxides (NO_x), and particulate matter (PM) highlight the key predictors influencing the emissions of each pollutant (Fig. 5). The analysis reveals notable differences in feature importance across the models, reflecting the distinct mechanisms governing their formation.

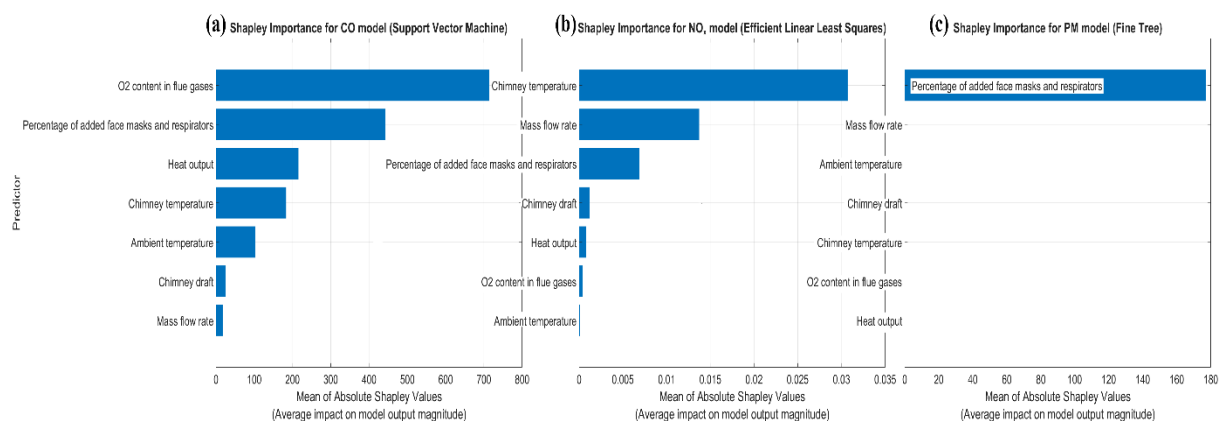


Fig. 5. Shapley importance for CO (a), NO_x (b), and PM (c) prediction using advanced machine learning models.

For the CO model, developed using a Support Vector Machine (SVM), the most influential predictor is O_2 content in flue gases, which aligns with fundamental combustion principles—higher oxygen levels improve combustion efficiency, thereby reducing CO emissions. The percentage of added face masks and respirators also shows a strong impact, suggesting a notable influence on CO formation, possibly through effects on combustion conditions or filtration efficiency. Other significant contributors include heat output and chimney temperature, indicating the role of energy output. Ambient temperature, chimney draft, and mass flow rate have relatively lower importance for that model.

In the NO_x model, using an efficient linear method, chimney temperature emerges as the most significant predictor. The mass flow rate follows as the second most influential factor, indicative of its role in the determination of fuel-air mixing and combustion conditions. The percentage of added face masks and respirators also contributes significantly, potentially modifying airflow and combustion characteristics. Surprisingly, ambient temperature and O₂ content in flue gases, which are typically crucial for NO_x formation, exhibit minimal influence in this model.

In the PM model, which uses a tree-based learning approach, the percentage of added face masks and respirators overwhelmingly dominates the feature importance. This finding suggests that changes in filtration or airflow caused by additional masks and respirators have a direct, substantial effect on PM levels. Other predictors, including mass flow rate, ambient temperature, chimney draft, chimney temperature, O₂ content in flue gases, and heat output, exhibit negligible influence in this model. Tree-based models, such as decision trees and random forests, are prone to high dependence on a single parameter. This is because these models work by dividing the data into branches based on the values of a single parameter in each node. Therefore, if one parameter is significantly more important than the others, the model may become overly dependent on it, thereby ignoring the others.

Variations in predictor significance across models indicate that effective emission control strategies must be tailored to specific pollutants. While optimizing oxygen levels can effectively reduce CO emissions, controlling NO_x and PM emissions may require adjustments in fuel composition, combustion temperature, and airflow dynamics. Furthermore, the observed limitations of tree-based models, particularly their sensitivity to dominant predictors, suggest that a hybrid modeling approach integrating multiple machine learning techniques could enhance predictive accuracy.

Conclusion

The conclusions of this study apply to pellets made from spruce sawdust with 0–50% (w/w) addition of crushed disposable masks and FFP2 respirators, combusted in an 18 kW retort-burner pellet boiler within the operating ranges listed in Table 1. The most meaningful next step is to expand the dataset across additional boiler types and to include quantified mask contamination and key fuel-property descriptors (e.g., moisture, ash, and elemental composition) as explicit predictors to improve model generalization.

This study demonstrated the feasibility of using multiple linear regression (MLR) and advanced machine learning (ML) models to predict CO, NO_x, and PM emissions from alternative fuel pellets containing varying amounts of crushed disposable face masks and FFP2 respirators. The results indicate that while MLR models provide an interpretable and straightforward approach to estimating emissions, advanced ML techniques significantly enhance predictive accuracy, particularly for CO and PM emissions. The best-performing ML models achieved high R² values during training, demonstrating near-perfect prediction accuracy for PM emissions. However, the validation phase revealed challenges with generalization, particularly for CO, suggesting the need for further refinement and a larger dataset to improve model robustness.

Compared to the basic multiple linear regression model, the advanced machine learning models demonstrate improved prediction accuracy, as evidenced by the RMSE values from the validation test on an independent dataset obtained during the combustion of a sample with a 10% addition of FFP2/face masks. For the CO model, the predictive performance was 31% higher based on RMSE values; for the NO_x model, it was 50.51% higher; and for the PM model, it was 94.05% higher.

Shapley importance analysis highlighted the dominant predictors for each emission type, confirming the crucial role of O₂ content in flue gases for CO, chimney temperature for NO_x, and the percentage of added face masks and respirators for PM emissions. The study also identified potential risks associated with excessive reliance on individual features in tree-based models, emphasizing the importance of careful feature selection and validation to avoid overfitting.

Overall, integrating ML models into combustion emission prediction offers a powerful tool for optimizing fuel blends and reducing environmental impact without requiring direct emissions measurements. Future research should explore real-time applications of ML-driven predictive control, investigate hybrid modeling approaches combining physics-based and data-driven methods, and refine models with larger and more diverse datasets to enhance generalization and reliability in practical applications.

References

- Adamczyk, W. P., Węcel, G., Kłajny, M., Kozołub, P., Klimanek, A., & Białecki, R. A. (2014). Modeling of particle transport and combustion phenomena in a large-scale circulating fluidized bed boiler using a hybrid Euler–Lagrange approach. *Particuology*, 16, 29–40. <https://doi.org/10.1016/J.PARTIC.2013.10.007>

- Baruah, D., Baruah, D. C., & Hazarika, M. K. (2017). Artificial neural network based modeling of biomass gasification in fixed bed downdraft gasifiers. *Biomass and Bioenergy*, 98, 264–271. <https://doi.org/10.1016/j.biombioe.2017.01.029>
- Belany, P., Hrabovsky, P., & Florkova, Z. (2024). Probability Calculation for Utilization of Photovoltaic Energy in Electric Vehicle Charging Stations. *Energies*, 17(5), 1073. <https://doi.org/10.3390/en17051073>
- Böhler, L., Fallmann, M., Görtler, G., Krail, J., Schittl, F., & Kozek, M. (2021). Emission limited model predictive control of a small-scale biomass furnace. *Applied Energy*, 285, 116414. <https://doi.org/10.1016/J.APENERGY.2020.116414>
- Boriouchkine, A., & Jämsä-Jounela, S. L. (2016). Simplification of a mechanistic model of biomass combustion for on-line computations. *Energies*, 9(9). <https://doi.org/10.3390/en9090735>
- Boriouchkine, A., Zakharov, A., & Jämsä-Jounela, S. L. (2012). Dynamic modeling of combustion in a BioGrate furnace: The effect of operation parameters on biomass firing. *Chemical Engineering Science*, 69(1), 669–678. <https://doi.org/10.1016/j.ces.2011.11.032>
- Čajová Kantová, N., Backa, A., Čaja, A., & Jandačka, J. (2025). Effect of fuel feeding cycle adjustments on emissions in spruce pellet combustion with the application of a four-tube electrostatic precipitator. *Scientific Reports*, 15, 25988. <https://doi.org/10.1038/s41598-025-11761-w>
- Čajová Kantová, N., Backa, A., Holubčík, M., Čaja, A., & Nosek, R. (2024a). Machine Learning Regression Model in MATLAB for Predicting Gas Emissions During the Pellet Combustion. In: *Smart Cities: Importance of Management and Innovations for Sustainable Development. Mobility IoT 2023. EAI/Springer Innovations in Communication and Computing*, 77–85. https://doi.org/10.1007/978-3-031-56533-5_6
- Čajová Kantová, N., Nosek, R., Čaja, A., Backa, A., & Vantúch, M. (2024b). Plastic medical waste: Energy use and environmental burden. [Plastový medicínsky odpad: Energetické využitie a environmentálna záťaž]. *Waste Forum*, 2, 138–143.
- Čajová Kantová, N., Nosek, R., Backa, A., Čaja, A., Jewiarz, M., & Mudryk, K. (2024c). Energy usage of spruce with waste face masks and spent coffee grounds as fuel in a pellet boiler. *Heliyon*, 10(15), e34802. [doi:10.1016/j.heliyon.2024.e34802](https://doi.org/10.1016/j.heliyon.2024.e34802)
- Čajová Kantová, N., Cibula, R., Szlek, A., Čaja, A., Nosek, R., & Belany, P. (2023). The energy assessment of COVID-19 medical waste as a potential fuel. *Energy Reports*, 9, 4995–5003. <https://doi.org/10.1016/J.EGYR.2023.04.018>
- Cao, H., Xin, Y., & Yuan, Q. (2016). Prediction of biochar yield from cattle manure pyrolysis via least squares support vector machine intelligent approach. *Bioresource Technology*, 202, 158–164. <https://doi.org/10.1016/J.BIORTECH.2015.12.024>
- Elmaz, F., Yücel, Ö., & Mutlu, A. Y. (2020). Predictive modeling of biomass gasification with machine learning-based regression methods. *Energy*, 191. <https://doi.org/10.1016/j.energy.2019.116541>
- EN ISO 17225-2: 2021. Solid biofuels — Fuel specifications and classes, Part 2: Graded wood pellets.
- Gauthier, G., & Avagianos, I. (2021). Pellets: Statistical Report 2021. Bioenergy Europe (European Pellet Council), Brussels.
- Grochowalski, J., Widuch, A., Śladek, S., Melka, B., Nowak, M., Klimanek, A., ... Adamczyk, W. (2023). Technique for reducing erosion in large-scale circulating fluidized bed units. *Powder Technology*, 426, 118651. <https://doi.org/10.1016/J.POWTECH.2023.118651>
- Ihme, M., Chung, W. T., & Mishra, A. A. (2022). Combustion machine learning: Principles, progress and prospects. *Progress in Energy and Combustion Science*, 91, 101010. <https://doi.org/10.1016/J.PECS.2022.101010>
- ISO 9096:2017: Stationary source emissions — Manual determination of mass concentration of particulate matter, Pub. L. No. 9096:2017 (2017).
- Jobson, J.D. (1991). Multiple Linear Regression. In: *Applied Multivariate Data Analysis. Springer Texts in Statistics*. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-0955-3_4
- Latosinska, J., Gawdzik, J., Honus, S., Orman, L. J., & Radek, N. (2023). Waste for building material production as a method of reducing environmental load and energy recovery. *Frontiers in Energy Research*, 11, 1279337. <https://doi.org/10.3389/fenrg.2023.1279337>
- Li, N., Lv, Y., & Hu, Y. (2023). Prediction of NO_x Emissions from a Coal-Fired Boiler Based on Convolutional Neural Networks with a Channel Attention Mechanism. *Energies*, 16(1), 76. <https://doi.org/10.3390/EN16010076>
- Lukáč, L., Kizek, J., Jablonský, G., & Karakash, Y. (2019). Defining the Mathematical Dependencies of NO_x and CO Emission Generation after Biomass Combustion in Low-Power Boiler. *Civil and Environmental Engineering Reports*, 29(3), 153–163. <https://doi.org/10.2478/ceer-2019-0031>
- Lukáč, L., Rimár, M., Variny, M., Kizek, J., Lukáč, P., Jablonský, G., Janošovský, J., & Fedák, M. (2020). Experimental Investigation of Primary De-NO_x Methods Application Effects on NO_x and CO Emissions from a Small-Scale Furnace. *Processes*, 8(8), 940. <https://doi.org/10.3390/pr8080940>

- Majewski, G., Orman, Ł. J., Telejko, M., Radek, N., Pietraszek, J., & Dudek, A. (2020). Assessment of thermal comfort in the intelligent buildings in view of providing high quality indoor environment. *Energies*, 13(8), 1973. <https://doi.org/10.3390/en13081973>
- Martinopoulos, G., Papakostas, K. T., & Papadopoulos, A. M. (2018, July 1). A comparative review of heating systems in EU countries, based on efficiency and fuel cost. *Renewable and Sustainable Energy Reviews*, 90, 687–699. <https://doi.org/10.1016/j.rser.2018.03.060>
- Morari, M., & Lee, J. H. (1999). Model predictive control: Past, present and future. In: *Computers and Chemical Engineering*, 23, 667–682. [https://doi.org/10.1016/S0098-1354\(98\)00301-9](https://doi.org/10.1016/S0098-1354(98)00301-9)
- Mutlu, A. Y., & Yucel, O. (2018). An artificial intelligence based approach to predicting syngas composition for downdraft biomass gasification. *Energy*, 165, 895–901. <https://doi.org/10.1016/J.ENERGY.2018.09.131>
- Nussbaumer, T. (2003). Combustion and Co-combustion of Biomass: Fundamentals, Technologies, and Primary Measures for Emission Reduction. *Energy and Fuels*, 17(6), 1510–1521. <https://doi.org/10.1021/EF030031Q>
- Orman, Ł. J., Majewski, G., Radek, N., & Pietraszek, J. (2022). Analysis of thermal comfort in intelligent and traditional buildings. *Energies*, 15(18), 6522. <https://doi.org/10.3390/en15186522>
- Ozgen, S., Caserini, S., Galante, S., Giugliano, M., Angelino, E., Marongiu, A., ... Morreale, C. (2014). Emission factors from small scale appliances burning wood and pellets. *Atmospheric Environment*, 94, 144–153. <https://doi.org/10.1016/J.ATMOENV.2014.05.032>
- Panchigar, D., Kar, K., Shukla, S., Mathew, R. M., Chadha, U., & Selvaraj, S. K. (2022). Machine learning-based CFD simulations: a review, models, open threats, and future tactics. *Neural Computing and Applications*, 34(24), 21677–21700. <https://doi.org/10.1007/s00521-022-07838-6>
- Permchart, W., & Kouprianov, V. I. (2004). Emission performance and combustion efficiency of a conical fluidized-bed combustor firing various biomass fuels. *Bioresource Technology*, 92(1), 83–91. <https://doi.org/10.1016/J.BIORTECH.2003.07.005>
- Porteiro, J., Míguez, J. L., Granada, E., & Moran, J. C. (2006). Mathematical modelling of the combustion of a single wood particle. *Fuel Processing Technology*, 87(2), 169–175. <https://doi.org/10.1016/j.fuproc.2005.08.012>
- Ross-Veitia, B. D., Palma-Ramírez, D., Arias-Gilart, R., Conde-García, R. E., Espinel-Hernández, A., Nuñez-Alvarez, J. R., ... Llosas-Albuérne, Y. E. (2024). Machine learning regression algorithms to predict emissions from steam boilers. *Heliyon*, 10(5), e26892. <https://doi.org/10.1016/J.HELİYON.2024.E26892>
- Van Der Lans, R. P., Pedersen, L. T., Jensen, A., Glarborg, P., & Dam-Johansen, K. (2000). Modelling and experiments of straw combustion in a grate furnace. *Biomass and Bioenergy*, 19(3), 199–208. [https://doi.org/10.1016/S0961-9534\(00\)00033-7](https://doi.org/10.1016/S0961-9534(00)00033-7)
- Xia, Y., Zhang, J., Tang, C., & Pan, W. (2023). Research and application of online monitoring of coal and biomass co-combustion and biomass combustion characteristics based on combustion flame. *Journal of the Energy Institute*, 108, 101191. <https://doi.org/10.1016/J.JOEL.2023.101191>
- Zadavec, T., Rajh, B., Kokalj, F., & Samec, N. (2020). CFD modelling of air staged combustion in a wood pellet boiler using the coupled modelling approach. *Thermal Science and Engineering Progress*, 20, 100715. <https://doi.org/10.1016/J.TSEP.2020.100715>
- Zha, Z., Ge, Z., Wang, S., Li, F., Ma, Y., Wu, Y., Zeng, M., Huang, Ch., Zhang, H. (2023). Biomass combustion behaviors based on online monitoring system: Morphological, thermodynamic and kinetic analysis. *Combustion and Flame*, 252, 112750. <https://doi.org/10.1016/J.COMBUSTFLAME.2023.112750>
- Zheng, Z.-H., Lin, X.-D., Yang, M., He, Z.-M., Bao, E., Zhang, H., & Tian, Z.-Y. (2020). Progress in the Application of Machine Learning in Combustion Studies. *ES Energy & Environment*, 1–14. <https://doi.org/10.30919/eseec8c795>