Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2025, 19(9), 452–468
https://doi.org/10.12913/22998624/208172
ISSN 2299-8624, License CC-BY 4.0

Received: 2025.06.15
Accepted: 2025.07.15
Published: 2025.08.01

Integrating experimental data and neural computation for emission forecasting in automotive systems

Magdalena Zimakowska-Laskowska[1*], Olga Anna Orynycz[2*], Ewa Kulesza[3], Jonas Matijošius[4], Karol Tucki[5], Antoni Świć[6]

[1] Environment Protection Centre, Motor Transport Institute, Poland
[2] Faculty of Engineering Management, Department of Production Management, Bialystok University of Technology, ul. Wiejska 45A, 15-351 Bialystok, Poland
[3] Faculty of Mechanical Engineering, Department of Mechanics and Applied Computer Science, Bialystok University of Technology, ul. Wiejska 45A, 15-351 Bialystok, Poland
[4] Mechanical Science Institute, Vilnius Gediminas Technical University, Plytinės Str. 25, LT-10105 Vilnius, Lithuania
[5] Institute of Mechanical Engineering, Department of Production Engineering, Warsaw University of Life Sciences, ul. Nowoursynowska 164, 02-787 Warsaw, Poland
[6] Department of Production Computerisation and Robotisation, Faculty of Mechanical Engineering, Lublin University of Technology, ul. Nadbystrzycka 36, 2018 Lublin, Poland
* Corresponding author's e-mail: magdalena.zimakowska-laskowska@its.waw.pl; o.orynycz@pb.edu.pl

ABSTRACT

This work presents an integrated approach combining experimental testing and mathematical modeling to analyze fuel consumption and pollutant emissions in a spark-ignition engine vehicle. Experimental data were obtained from chassis dynamometer tests under the WLTP driving cycle, including time series of vehicle speed, energy consumption, and CO_2, CO, THC, and other compounds emissions. Two classes of artificial neural networks were implemented to capture the complex, nonlinear relationships between driving dynamics and emission profiles: Multi-layer perceptrons (MLP) and self-associative neural networks (SANN). These models were trained on real-world time series data to predict vehicle speed and energy consumption as functions of emission parameters and vice versa. The models demonstrated high accuracy, especially in the validation phase, confirming their potential for forecasting and environmental performance assessment. The neural network models underwent training, validation, and testing processes, allowing for the assessment of their effectiveness in predicting energy consumption under various system operating scenarios. The results demonstrated high prediction accuracy, confirming the usefulness of ANN as a tool for analyzing complex relationships between emissions and energy efficiency. The best-performing model achieved a mean absolute error (MAE) of 0.034 MJ/km and a coefficient of determination (R^2) of 0.91 for energy consumption prediction. The study developed models identifying the relationships between emission parameters and energy consumption characteristics, enabling precise modeling of combustion processes. The input data included key emission indicators such as carbon monoxide (CO), carbon dioxide (CO_2), and hydrocarbons (HC, NMHC and CH_4), as well as operational parameters of energy systems. Additionally, the observed patterns in energy use were interpreted through a physical lens, considering the thermodynamics and chemical kinetics of combustion processes under different driving conditions. This hybrid methodology – combining data-driven AI with domain-specific physical insight – provides a robust framework for predicting the environmental impacts of internal combustion engines and optimizing their operation. The proposed approach applies to broader engineering contexts, including emission control strategy design, digital twin development for powertrains, and intelligent vehicle energy management systems. The proposed approach represents a significant step toward leveraging modern artificial intelligence methods to improve energy efficiency and develop emission reduction strategies through combustion condition optimization. The obtained results can serve as a foundation for further refining industrial processes in the context of sustainable development and environmental protection.

Keywords: emission modeling, energy efficiency, combustion process optimization, sustainable development, energy consumption prediction, artificial neural networks

INTRODUCTION

Reducing emissions from the transport sector is one of the key challenges in the context of global efforts to protect the climate and sustainable development. Vehicles with combustion engines, especially those used in urban environments, emit significant amounts of greenhouse gases and harmful compounds such as nitrogen oxides (NOx), carbon monoxide (CO) and hydrocarbons (THC). Traditional emission estimation models – such as the widely used COPERT – are based on simplified relationships between average speed and emissions, which do not allow for capturing the real driving dynamics, including acceleration, braking or standstill phases. As shown by Lejri and Samaras et al., this approach may lead to significant underestimation of emissions, up to 25% in congested conditions [1, 2]. In response to these limitations, increasing attention is being paid to artificial intelligence models, in particular neural networks. Their application in energy-related socioeconomic analysis was demonstrated by Kulisz et al. [3], who used neural networks with Bayesian regularization to predict household heating inadequacy across Europe, achieving high accuracy. These models, trained on real measurement data, can represent complex, nonlinear relationships between engine performance parameters, traffic conditions, and emission levels. A similar approach has been used to analyze the impact of road surfaces on vehicle dynamics, where neural networks have shown high efficiency in classifying road surface types based on accelerometer data [4].

The classical approach to creating and training neural networks is described in detail in Haykin's work [5]. This approach fits into the paradigm of deep machine learning, described by Goodfellow [6], among others. This paper presents a hybrid approach integrating data from tests conducted on a chassis dynamometer according to the WLTP procedure (worldwide harmonized light vehicles test procedure) with calculations based on artificial neural networks (MLP, SANN), aimed at improving the accuracy of emission and energy consumption forecasts. A similar approach, combining chassis dynamometer data with computational models, was used in the work of Zimakowska-Laskowska, where the synergistic potential of experimental and numerical tools in the analysis of vehicle exhaust emissions was demonstrated [7, 8].

Therefore, there is a growing need for advanced analytical tools to accurately capture the complex, nonlinear dependencies between engine operation, vehicle dynamics, and emissions. When trained on real measurement data, artificial neural networks offer high prediction accuracy even under transient operating conditions. This study addresses this gap by integrating experimental WLTP-based emission data with predictive models to improve the reliability of emission and energy consumption estimation. The proposed approach contributes to developing digital twins and advanced emission control strategies in modern transport systems. The study aims to create and validate a predictive framework combining experimental emission data and artificial neural networks to accurately estimate energy consumption and exhaust emissions under real-world driving conditions.

BACKGROUND

According to the current state of knowledge, artificial neural networks are highly effective in predicting vehicle emissions. MLP (Multi- Layer Perceptron) models and their deep counterparts, such as DNN, allow for accurate mapping of CO_2, NOx, CO and THC emissions based on engine operating parameters. In Seo studies achieved correlation coefficients of up to 0.98 for CO and 0.96 for CO_2, with relative errors below 1% [9]. Also Fu et al. and Hashemi confirmed the high precision of neural networks in applications to SI and CI engines [10, 11]. The effectiveness of ANN was also confirmed in the prediction of emissions from HVO-fueled diesel engines, where, despite the challenges related to data validation, high accuracy of CO and NOx prediction was achieved in various operating regimes [12].

The choice of input data is crucial – the most commonly used are: vehicle speed, torque, air/fuel ratio, coolant temperature, and mass air flow. The selection of these parameters can be improved by Pearson correlation analysis and p-significance tests, which improves the efficiency and accuracy of the models.

In relation to emissions in real conditions (e.g. cold start), ANNs also show great usefulness, although the prediction accuracy for some components – especially THC – may be lower. LSTM models, belonging to the class of recurrent networks, also show high efficiency in

forecasting variable energy phenomena, which was confirmed, among others, in the analysis of energy production in PV systems using real data. Models based on time-aware architectures (e.g. LSTM, SANN) cope better with such dynamic phenomena. A detailed review of the applications of deep neural networks in time series analysis was presented by Kong [13]. An example is the work of Shin in which a DNN model optimized by the Bayesian method achieved $R^2 = 0.9675$ in the forecast of NOx emissions based on WLTP data [14]. This approach also finds application in modeling long-term CO_2 emissions at a national scale, where deep neural networks – especially LSTM – have shown superior performance in analyzing high-variability time series.

An analysis of the possibility of using neural networks to determine the parameters of the chemical composition of exhaust gases as a function of engine performance parameters obtained from the on-board diagnostics system such as crankshaft speed and engine load index was presented in [15]. For the purpose of building a neural network model, preliminary studies were carried out in non-urban traffic (high-speed route). On the basis of the data obtained, processes of learning neural network structures with approximate properties with backward propagation of errors were carried out. Subsequently, tests were performed on the operational parameters of the vehicle and the chemical composition of exhaust gases in urban traffic.

Artificial neural network models for forecasting the combustion and emission characteristics of ethanol/gasoline DFSI engines with combined injection strategy were presented in [16]. Models of artificial neural networks (ANN) were developed to reflect the performance of DFSI engines. The regression values were within the range of 0.9387–0.9962, and the mean square relative errors were within the range of 0.000184–0.03935 between the ANN predicted and experimentally measured results.

A biofuel-powered study with deep learning neural networks and dragonfly algorithm: optimizing CRDI engine performance with ZnO nanoparticles and cotton seed methyl ester was presented in [16]. The engine combustion characters are modelled using deep learning neural networks (DNN) and single-layered neural networks (ANN). Deep learning models accurately predicted key engine parameters like heat release rate (HRR) and in-cylinder pressure (ICP), achieving $R2 > 0.95$. The engine emissions are within an acceptable range, and the BTE is within the range of 20% to 32% as a result of the engine performance optimization.

An evaluation of machine learning algorithms on hydrogen boosted homogeneous charge compression ignition engine operation for performance and emission prediction was presented in [17]. The dataset included three input parameters namely hydrogen energy share (HES), equivalence ratio and injection timing, five output parameters like brake thermal efficiency (BTE), NOx, smoke, hydrocarbon (HC), CO. The total of 26 machine learning algorithms were trained and tested. The efficient machine learning model was identified by synthesized evaluation of MSE, root mean square error (RMSE), R-squared (R^2) and MAE values. Among the algorithms considered, Matern 5/2 GPR, Wide Neural Network and Fine Tree algorithm were excellent for predicting the BTE with the R^2 value of 0.9999, 0.9985, and 0.9961.

Artificial neural network and fuzzy expert system comparison for prediction of performance and emission parameters on a gasoline engine was presented in [18]. The study presented used ANN and a fuzzy expert system (FES) to model a gasoline engine in order to predict engine power, torque, specific fuel consumption, and hydrocarbon emissions. As a result, it has been shown that developed ANN and FES can be used reliably in automotive industry and engineering instead of experimental work.

An artificial neural network model to predict efficiency and emissions of a gasoline engine was presented in [19]. In the study, an ANN model was built to predict three types of indicators (power, emissions, and combustion phasing) together, including 50% combustion crank angle (CA50), CO, unburned hydrocarbons (UHC), NOx, indicated mean effective pressure (IMEP), and indicated thermal efficiency (ITE). The goal of this work was to verify that only one machine learning model can combine power, emissions, and phase metrics together for prediction. The predicted results showed that all coefficients of determination (R^2) were larger than 0.97 with a relatively small RMSE, indicating that it is possible to build a predictive model with three types of parameters (power, emissions, phase) as outputs based on only one ANN model.

Modeling of CO emissions from traffic vehicles using artificial neural networkswas presented in [20]. In the paper was presented a hybrid model

based on data mining and GIS models designed to predict vehicular CO emitted from traffic on the New Klang Valley Expressway, Malaysia. The model was developed using six traffic CO predictors: number of vehicles, number of heavy vehicles, number of motorbikes, temperature, wind speed and a digital surface model. The network architecture and its hyperparameters were optimized through a grid search approach. The traffic CO concentrations were observed at 15-min intervals on weekends and weekdays, four times per day. The results showed that the developed model had achieved validation accuracy of 80.6 %.

Predicting Fuel Consumption and Emissions Using GPS-Based Machine Learning Models for Gasoline and Diesel Vehicles was presented in [21]. In the study was investigated the use of real-world driving data from gasoline and diesel vehicles to model fuel consumption and exhaust emissions (CO_2 and NO_X). The results demonstrate high predictive accuracy, with the ensemble bagged model consistently outperforming the decision tree model across all datasets.

Equally promising results were obtained in the field of fuel consumption forecasting. ANN, SVM models and ensemble ML methods (random forest, gradient boosting) allow for errors of around 1.7% and determination coefficients above 0.90. The classic SVM method was developed by Cortes and has been the foundation of many machine learning algorithms for years [22]. Data on vehicle speed, vehicle weight, road gradient, as well as weather conditions and load are important here.

Integrating measurement data from the WLTP procedure with neural models allows for a significant increase in the accuracy of emission and energy consumption prediction – also in transient conditions. As demonstrated by Dini this approach is also applicable in monitoring the aging of drive systems and early detection of anomalies [23]. Integrating measurement data from the WLTP procedure with neural models allows for a significant increase in the accuracy of emission and energy consumption prediction – also in transient conditions, which was confirmed, among others, in studies on emission prediction in various transport scenarios.

In summary, the literature review confirms the validity of using ANNs for predicting vehicle emissions and energy efficiency. Their effectiveness in predicting emission time series values is also confirmed by the work of Tlelo-Cuautle [24]. The integration of real data with machine learning algorithms is the direction of development of modern tools supporting environmental management, development of digital twins of drive systems and intelligent transport systems.

METHODOLOGY

Research site and data acquisition

The tests were carried out on a passenger car meeting the Euro 6 emission standard, equipped with a spark-ignition (SI) engine. The vehicle had a typical B-segment curb weight of around 1200 kg, front-wheel drive and a six-speed gearbox. The vehicle was not modified for the tests, which ensured that the results were consistent with real-world operating conditions. The tested vehicle was new and had very low mileage at the time of testing, ensuring minimal influence of component wear on emission results.

The measurements were taken in laboratory conditions, using a chassis dynamometer and devices recording exhaust emissions in real time. Exhaust emissions were measured using a set of dual-range gas analyzers (AVL AMA i60 R2 and CEB II), compliant with WLTP standards. The system allowed for the measurement of CO, CO_2, THC, CH_4, NOx (including NO and NO_2), and N_2O, with high temporal resolution. The equipment enabled both pre- and post-catalyst measurements as well as EGR rate estimation. The WLTP (Worldwide Harmonized Light Vehicles Test Procedure), which is currently the homologation standard for vehicles in the European Union. Only data from the urban phase (low speed) and suburban (medium speed), which corresponds to typical operating conditions in an urbanized environment.

The recorded parameters included vehicle speed (v, in km/h), instantaneous energy consumption (EC, in MJ/km), and concentrations of gaseous pollutants such as CO_2, CO, total THC, NOx, CH_4, and particulate matter (PM).

The measurements were recorded at a frequency of 1 Hz, which allowed obtaining accurate time series for each parameter. The raw data were then subjected to preliminary cleaning (removal of missing data, anomaly filtration) and normalization – each variable was rescaled to the range [0,1] to ensure the stability of the neural model training process. This type of normalization is a standard element of preparing input data for machine learning. In the context of models

based on kernel functions, such as SVM, data rescale is particularly important for the correctness of the model.

The data set was divided into three subsets:

- training set (70%) – used to train models,
- validation set (15%) – used to optimize hyperparameters and prevent overfitting,
- test set (15%) – enabling an independent assessment of prediction accuracy.

The final dataset included 17 complete observation cycles, which were divided into training (12), validation (2), and test (3) subsets.

This data structure provides appropriate conditions for building and evaluating predictive models based on artificial neural networks, enabling the estimation of both energy consumption and emissions of individual exhaust gas components.

The preliminary correlation analysis between CO_2 emissions and energy consumption showed a low correlation coefficient ($r \approx 0.22$), which suggests a nonlinear nature of the relationship and justifies the use of nonlinear methods such as neural networks. The use of correlation and statistical analysis as a preliminary modeling step has been extensively discussed by Hastie.

Mathematical tools used

Two types of artificial neural networks were used to model pollutant emissions and energy consumption: a multilayer perceptron network (MLP) and a simple artificial neural network. Neural network (SANN). Both models were trained on the experimental dataset obtained from tests compliant with the WLTC (worldwide harmonized light duty vehicles test cycle). The data included instantaneous speed, torque, mass fuel consumption, coolant temperature and measured CO, CO_2 and HC emission values.

In order to map the relationships between vehicle operating parameters and energy consumption and emissions, predictive models based on machine learning methods have been developed. Advanced alternative approaches to classical ML models also appear in the literature, such as high-dimensional model representation (HDMR), which enables effective mapping of complex relationships without the need for full numerical calculations. Three classes of algorithms were used: artificial neural networks (MLP and SANN) and SVM. This section presents the architecture of the models used, the method of data preparation and the metrics for assessing the effectiveness of prediction. This approach allows for the analysis of complex and nonlinear relationships characteristic of the conditions of real vehicle operation.

The selection of MLP, SANN, and SVM models was based on their proven effectiveness in modeling complex, nonlinear relationships typical for vehicle emissions and energy consumption. MLP networks are widely used due to their universal approximation capability and robustness in regression tasks involving multidimensional input spaces. SANNs, as simplified network structures, offer faster training times and lower computational costs, making them suitable for initial testing and baseline comparisons. SVM with the RBF kernel was chosen as a classical yet powerful algorithm capable of handling small datasets with high dimensionality while maintaining good generalization performance. These models have been frequently applied in related studies, demonstrating high accuracy in emission prediction tasks. Their selection in this study was also motivated by the desire to compare the performance of both neural and kernel-based approaches in the same experimental context.

The machine learning process. which usually takes place in several stages: data preparation. model selection. model training. performance evaluation and optimization. All of these stages can be described using mathematical equations and concepts.

According to literature data, it can be described as follows:

Data preparation

In the machine learning process, it is crucial to prepare the appropriate data. which will be used to train the model. Usually, we have a data set D={(x1.y1).(x2.y2).….(xn.yn)}. Where xi are features (input vectors), and yi are output values, e.g. classes in the case of classification. The goal of modeling is to identify a function that maps the relationship between the feature space x and the predicted values y.

Model selection

A machine learning model can be described as a function that maps input data x to its corresponding predicted value y. Mathematically, the model represents a function f: X→Y. where X is the feature space (all possible inputs). and Y is

the value space. Depending on the problem, it can be either regression (where y is a real number) or classification (where y is a category).

The model equation in the case of linear regression is expressed by the formula (1):

$$y = w^T x + by \quad (1)$$

where: w is the weight vector, x is the feature vector, b is the intercept, and y is the value predicted by the model.

Training the model

The goal of training a model is to find parameters (e.g. weights in linear models) that minimize the prediction error. In the case of classification, the most common method is to minimize the loss function, which calculates the difference between the model predictions and the actual y values. For linear regression, the standard loss function is the MSE (2) [25]:

$$L(w,b) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (w^T x_i + b)\right)^2 \quad (2)$$

where: L is the loss function, w are the weight vectors, xi are the features, and yi are the actual values of.

During training we optimize the model parameters (e.g. wib) to minimize the loss function. The most commonly used method is gradient descent (3) [26]:

$$w \longleftarrow w - \eta \nabla_w L(w,b) \quad (3)$$

where: η is the learning rate and $\nabla_w L(w, b)$ is the gradient of the loss function with respect to the weights.

Performance evaluation

Accuracy are often used for classification (4) [26]:

$$accuracy = \frac{1}{n}\sum_{i=1}^{n} II(f(x_i) = y_i) \quad (4)$$

where: I is an indicator function, which is 1 if the model prediction is consistent with the true value of and 0 otherwise. (Heaviside unity).

In the case of regression. you can use the MAE or the MSE (5, 6) [25]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (5)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}i)^2 \quad (6)$$

where: yi – the actual values, and $\hat{y}_i$ – the model predictions.

Model optimization

After evaluating the model. if its performance is not satisfactory. various optimization techniques are used. such as regularization. changing the model architecture. or using advanced algorithms. e.g. global optimization algorithms. decision trees. neural networks.

Support vector machines (SVM)

Modeling process for SVM with RBF kernel (Radial Base Function) can be described in several stages. taking into account specific parameters. e.g.: C = 7C = 7C = 7. ϵ = 0.1. and γ = 1.75. Below will be describe the details of the mathematics and the RBF kernel function. as well as how these parameters affect the model.

Data preparation

As in every case in machine learning. the first step is to prepare the data in the form of pairs (x_i, y_i) where $x_i \in R^d$ are feature vectors. a $y_i \in \{-1.1\}$ are the classification values (assuming we have a binary classification problem). The dataset D = $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is then used to train the SVM model.

Kernel function definition

SVM is an algorithm that tries to find the best boundary (hyperplane) separating classes in feature space. The key element in SVM is the kernel function. which allows to transform the input data into higher dimensionality. in which it is easier to find a linear separating boundary.

In the case of the RBF kernel, the kernel function is given by (7) [27]:

$$K(x_i, x_j) = exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \quad (7)$$

where: γ is a parameter that controls the ”stretch” of the kernel. The value of γ affects how much the data will be ”smeared” in feature

space. The value of $\gamma = 1.75$ that you provided means that the kernel will be relatively ”wide”. This can make the model more sensitive to local changes in the data.

Optimization goal – Maximize margin

The goal of the SVM algorithm is to find a hyperplane that maximizes the margin (the distance between the closest data points from both classes). From a mathematical perspective, the optimization objective in SVM classification is to find a hyperplane that maximizes the margin between classes. This is equivalent to solving the following problem (8) [27]:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{8}$$

While maintaining the following conditions (9):

$$y_i(w^T x_i + b) \geq 1, \textit{ for all } i = 1, 2, \ldots, n \tag{9}$$

where: w – the weight vector, b – the intercept, y_i – the classification value.

Loss function and regularization

In SVM the loss function is combined with regularization. We introduce the parameter CCC. which controls the trade-off between maximizing the margin and allowing for classification errors. The value C = 7C = 7C = 7. that you provided. means that the model is relatively intolerant of classification errors – we want all points to be correctly classified. but not at any cost.

The loss function of SVM with regularization has the form (10) [25]:

$$L(w, b, \xi_i) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \tag{10}$$

where: ξi are truncated variables that represent classification errors (distances of points that are misclassified or fall in the margin zone). The value of C controls how large these errors can be.

Optimization

Optimization in the case of SVM is done using the dual method (in Lagrangian space). Using the RBF kernel, the optimization problem takes the form (11) [26]:

$$\min_{\alpha} \left(\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right) \tag{11}$$

Lagrange variables and $K(x_i,x_j)$ is the kernel function. in this case RBF. The variable αi is responsible for the weights assigned to the training data. which determine which points have the greatest influence on the definition of the hyperplane (so-called ”support vectors”).

Performance evaluation

The model, its performance needs to be evaluated. SVM, like other classification models, can be evaluated using metrics such as accuracy. precision. recall. and F1 – score. In case of regression problem, other metrics are used, such as MSE.

Final result

After optimizing the model, we use it to predict the class for new points. If $f(x) = w^T x + b$ decision based on a decision function, then the classification is made based on the sign of this function (12):

$$\hat{y} = sign\big(f(x)\big) \tag{12}$$

where: $\hat{y}$ is the predicted value of af(x)f(x)f(x) is the result of the decision function.

Parameter summary

- C = 7C = 7C = 7 – the model is less tolerant to classification errors and tries to best fit the training data.
- $\epsilon = 0.1$ – the epsilon parameter in SVM for regression controls the margin of allowable errors. and in the case of classification, ϵ can refer to the tolerance in evaluating the results.
- $\gamma = 1.75$ – the γ parameter in the RBF kernel controls the shape of the kernel function, deciding how much to “stretch” the influence of the training points. A higher γ makes the model more sensitive to local structures in the data, and a low γ makes the decision boundary smoother (Table 1).

Due to the limited sample size (n = 17), the obtained results should be considered preliminary and illustrative. They constitute a technical confirmation of the correctness of the adopted methodology, but do not yet allow for a full assessment

of its effectiveness. In further stages, it is planned to expand the data set and re-validate the models using more diverse test scenarios.

The SVM model with the RBF kernel showed a very good fit to the training data. At the same time, significant differences in the MSE values and the correlation coefficient between the training and test sets indicate the presence of overfitting. The high number of support vectors in relation to the number of observations (9 out of 17) confirms the relatively high complexity of the model in the context of the limited sample. Improving the generalization ability requires both increasing the number of data and optimizing the hyperparameters, including the selection of the kernel and validation procedures.

RESULTS AND DISCUSSION

This chapter presents the results of the applied predictive models (MLP, SANN, SVM) and their evaluation in terms of the effectiveness of pollutant emission mapping and specific energy consumption. The accuracy was verified based on classical regression metrics: MSE, correlation coefficient (R) and standard deviation ratio (SD ratio). A similar approach to the evaluation of the effectiveness of ML models was also used in the analysis of production processes, including the prediction of surface roughness in milling based on sensor signals. A novel application of image analysis combined with physical property evaluation in agricultural materials was recently demonstrated by Gierz et al. [25], indicating the potential of cross-domain ML applications. The effectiveness of ML in the analysis of measurement data was also noticed in material diagnostics, an example of which is the application of deep learning methods to ultrasonic tomography. A similar approach was also implemented in the monitoring of hydrotechnical infrastructure, where ML models supported the analysis of impedance tomography images to detect leaks in flood embankments. In addition to industrial and environmental applications, deep learning models are also gaining recognition in road transport, including in adaptive fuel consumption modeling using meta-learning and on-board data. In agricultural machinery, Al-Sammarraie et al. [26] successfully applied machine learning for predicting power take-off performance, demonstrating the model's flexibility beyond traditional transportation contexts. The importance of ML models for fuel consumption forecasting is also confirmed by analyses conducted in the heavy goods vehicle sector, where classical algorithms were used to estimate average fuel consumption based on operational data.

Additionally, a spatial analysis of the relationship between driving speed, energy consumption and emissions of individual exhaust components was performed, using models to generate predictive surfaces. These results were compared with literature conclusions, enabling the assessment of the usefulness of selected algorithms in the context of modeling operational and environmental processes.

In summary, the research methodology follows a structured pipeline: real-time emission data acquisition in controlled conditions, data preprocessing and normalisation, selection and training of machine learning models (MLP, SANN, SVM), and evaluation of predictive performance using standard statistical metrics. Established standards for vehicle emission testing and machine learning were implemented at each stage. This approach ensures transparency, repeatability, and the potential for extension to broader datasets and operational scenarios.

Model prediction performance (MLP, SANN, SVM)

Comparison of the results obtained by the applied AI models reveals significant differences in their predictive performance. The 6-10-1 MLP

Table 1. Assessment of the quality of the SVM model fit for energy consumption prediction (EC [MJ/km])

Metrics	Training set	Test set	The whole (overall)	Comment
Number of samples	12	5	17	Small test set – Limited reliability
Number of support vectors	9 (6 limited)	-	-	High SV number → possible overfitting
MSE	6.769	16.180	9.537	A significant difference between train and test – overfitting
Correlation coefficient (R)	0.931	0.816	0.905	Good fit but poorer generalization
SD ratio	0.377	0.611	0.435	Larger relative error in the test set

model achieved an almost perfect fit to the training data (R = 0.998) with a very low MSE (approx. 2.21 MJ²/km²). At the same time, a significantly higher error in the test set (MSE > 16 MJ²/km²) and a decrease in R to ~0.85 indicate strong overfitting. This is a typical effect for networks trained on small data sets and requires further optimization using regularization or cross-validation. This model was also characterized by a low SD ratio (0.389), which confirms the stability of its predictions.

In turn, the SANN model, despite slightly lower correlation coefficients, proved particularly effective in predicting dynamic variables – such as speed or instantaneous changes in energy consumption (EC) – thanks to its ability to process time series. It also reproduced well the shape of the curves in transient conditions, such as acceleration and braking phases in the WLTP cycle.

The SVM model, despite the high R = 0.930 in the training set, achieved a significantly lower test result (R = 0.824), and at the same time a higher MSE = 15.75 MJ²/km² and SD ratio = 0.610. This indicates the presence of overfitting, especially at a high value of the regularization parameter C = 7.000. The model would require further optimization or the use of dimensionality reduction techniques. Table 2 presents the detailed results of the SVM models as a function of the variable gamma parameter at C = 7.

Due to the very limited size of the dataset (n = 17), the obtained results should be considered preliminary. Such a small sample size may affect the instability of metric estimation and susceptibility of models to overfitting, even with seemingly optimal hyperparameter settings.

Table 2 shows the effect of the gamma parameter on the accuracy of the SVM model at a constant value of C = 7. With the increase of gamma, a moderate increase in the MSE error and a decrease in the correlation coefficient R are noticeable. The best compromise between accuracy and stability was achieved for gamma = 0.175, for which the lowest test error (MSE = 15.75) and the highest correlation coefficient (R = 0.824) were obtained. The SD ratio also indicates the optimal generalization of the model with this configuration. Gamma values above 0.200 resulted in a decrease in the prediction accuracy and an increase in the susceptibility to overfitting.

The obtained results indicate an almost perfect fit of the model to the training data, with significantly weaker performance in the test set. This suggests strong overfitting, typical for neural networks trained on small data sets. It is necessary to use regularization and cross-validation mechanisms.

Machine learning evaluation – SVM model for energy consumption prediction (EC [MJ/km])

For the SVM model with RBF kernel [27], a sample size of 17 observations was used, of which 12 were used for training and 5 for testing. The number of support vectors was 9, including 6 constrained ones, which means that a significant part of the sample was used to determine decision boundaries, which may indicate model complexity and the risk of overfitting.

The MSE was 6.769 for the training set, 16.180 for the test set, and 9.537 for the entire data set. The significant difference between the training and test errors suggests the presence of overfitting, i.e. excessive fitting of the model to the training data at the expense of generalization ability.

The correlation coefficient (R) also confirms this trend – in the training set the value was 0.931, while in the test set it was 0.816. The overall value was 0.905, which indicates a good quality of fit, however the decrease in R in the test set may indicate limited effectiveness of the model in predicting new cases.

The SD ratio reached 0.377 for the training set, 0.611 for the test set and 0.435 overall. The low value in the training set confirms the good agreement of the model with the training data, while the higher value in the test set additionally confirms the weaker predictive ability.

In summary, the SVM model showed a high quality of fit to the training data, but relatively poor generalization. The high number of support vectors relative to the number of observations and the significant difference between Train and Test metrics suggest the need to optimize hyperparameters and increase the size of the dataset (Table 3).

The obtained results indicate a very good fit of the SANN model to the training data. Although the correlation coefficient in the test set is high (R = 0.934), the significantly higher value of the MSE error relative to the training set (22.94 vs. 11.64 MJ²/km²) may indicate prediction instability or the presence of outliers. This difference suggests a limited ability of the model to generalize, which may result from a small number of observations and the need for further optimization of the network structure.

Table 2. Comparison of the performance of SVM models for different gamma values (C = 7)

Gamma	MSE (Train)	MSE (Test)	R (Test)	SD ratio (Test)
0.140	8.548	17,658	0.822	0.638
0.150	8.344	17,377	0.821	0.634
0.167	7.097	15,751	0.824	0.610
0.175	6.936	15.648	0.824	0.607
0.180	6.936	15.679	0.824	0.607
0.200	6.818	15,771	0.822	0.607
0.250	6.522	15.931	0.818	0.604
0.300	6.326	16.509	0.815	0.603

Table 3. Assessment of the quality of the SVM model fit for energy consumption prediction (EC [MJ/km])

Metrics	Training set	Test set	The whole (overall)	Comment
Number of samples	12	5	17	Small test set – limited reliability
Number of support vectors	9 (6 limited)	-	-	High SV number → possible overfitting
MSE	6.769	16.180	9.537	A significant difference between train and test – overfitting
Correlation coefficient (R)	0.931	0.816	0.905	Good fit but poorer generalization
SD ratio	0.377	0.611	0.435	Larger relative error in the test set

Error and correlation analysis

The analysis of residual error plots showed that the MLP and SANN models best reproduce the data in the speed range of 40–70 km/h, which corresponds to the optimal energy consumption. In the very low speed range (< 10 km/h), an increase in prediction errors was observed, probably related to the greater randomness of driver behavior and reduced efficiency of the drive system during stops and starts.

This trend is illustrated in Figure 1, which shows a clear inverse relationship between specific EC and vehicle speed.

The graph shows a clear downward trend: at low speeds, higher specific energy consumption is observed. This may be due to intensive acceleration phases and unfavourable operating conditions of the drive system at low speeds.

One of the key areas of analysis was the relationship between specific energy consumption, driving speed and CO_2 emissions. MLP and SANN models made it possible to reproduce this relationship in a spatial system, revealing areas of optimal drive system operation (Figure 2).

In the center of the graph, the minimum CO_2 emissions are observed, corresponding to the range of moderate speed and optimal energy consumption. Both low speeds with high EC and high speeds with increasing EC result in a significant increase in emissions, reflecting the typical trade-off between dynamics and combustion efficiency.

In order to identify areas of intense CO emission, an analysis of the spatial relationship between EC, speed and predicted emission level was performed (Figure 3). The models enabled the identification of areas where the drive system shows the highest losses in the context of fuel oxidation (Figure 3).

CO emissions show two characteristic growth zones: at low speed and high specific energy consumption (starting and driving conditions) and at higher speeds and intensive loads. Minimum emission values are observed in the moderate speed and medium EC range, suggesting the existence of an optimum engine operating range in terms of carbon monoxide emissions.

In order to assess the conditions favoring the emission of THC, their dependence on specific energy consumption and driving speed was analyzed. The models made it possible to generate the spatial distribution of emissions, allowing to indicate the areas with the highest emissions (Figure 4).

THC emissions are at their highest in conditions of low speed and high specific energy consumption – typical of starting, stopping and running cold. A second increase in emissions occurs at very high EC, which may be related to overloading the drive system. Minimum emissions are observed in the middle range of EC and speed, which indicates the

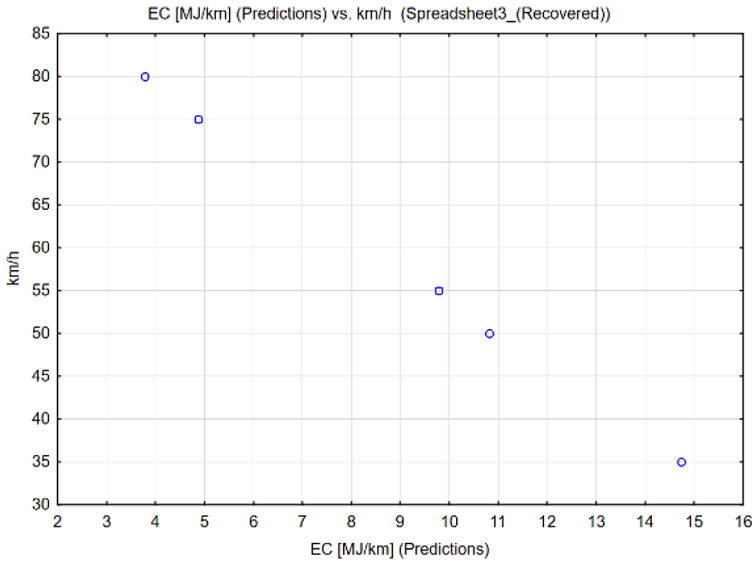

Figure 1. Relationship between energy consumption and vehicle speed (predicted data)

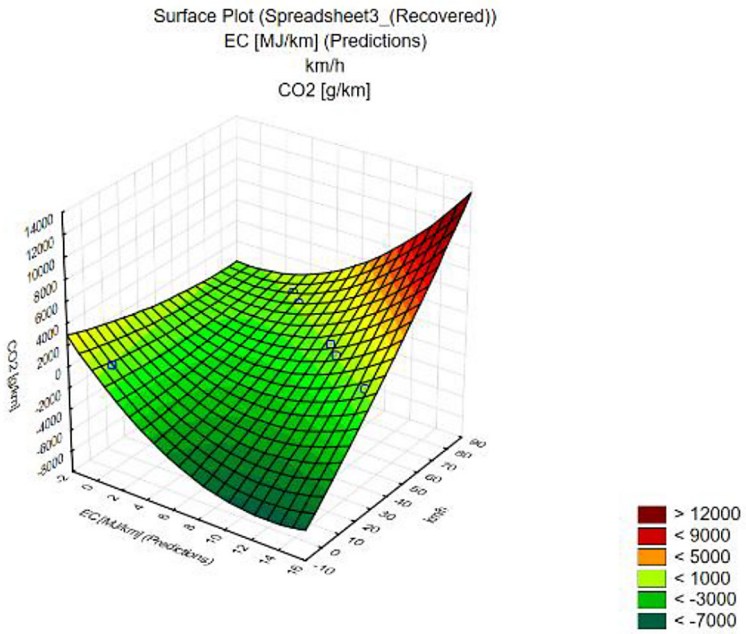

Figure 2. 3D relationship surface between energy consumption (EC), vehicle speed and CO_2 emissions (ML model predictions)

existence of an effective operating regime limiting the emission of unburned hydrocarbons.

Methane (CH_4), as a representative of emissions characteristic of gaseous fuels, shows a strong dependence on vehicle operating conditions. Based on ML models, spatial distributions of CH_4 emissions were determined as a function of speed and specific energy consumption (Figure 5).

Methane emissions increase significantly at low speed and high EC, which may be the result of incomplete combustion during start-up and insufficient catalyst temperature. A small increase in emissions is also observed at extremely high EC values, which may be related to engine overload. The minimum CH_4 emissions occur in the moderate speed range and average energy consumption – similar to the observations for THC.

For a more complete characterization of hydrocarbon emissions, an analysis of the non-methane component (NMHC) was also performed.

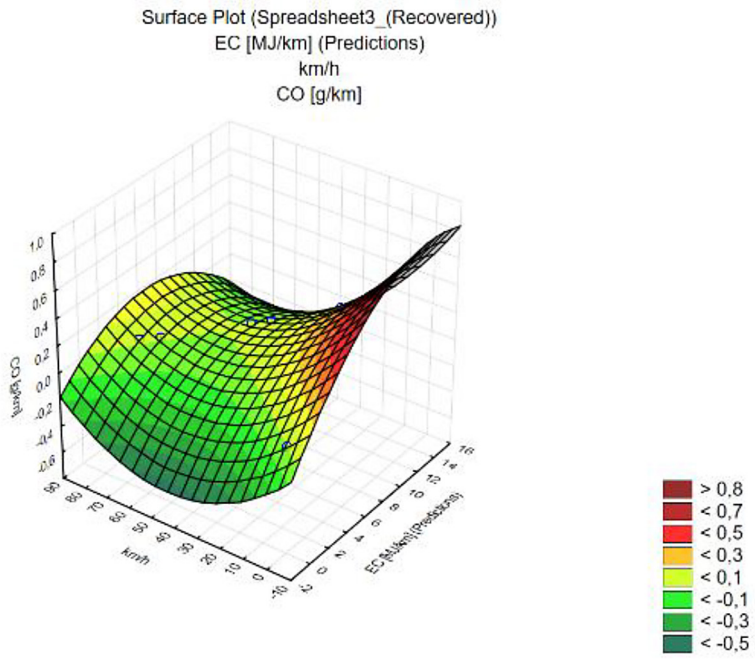

Figure 3. Spatial relationship between specific EC, vehicle speed and CO emissions (ML model predictions)

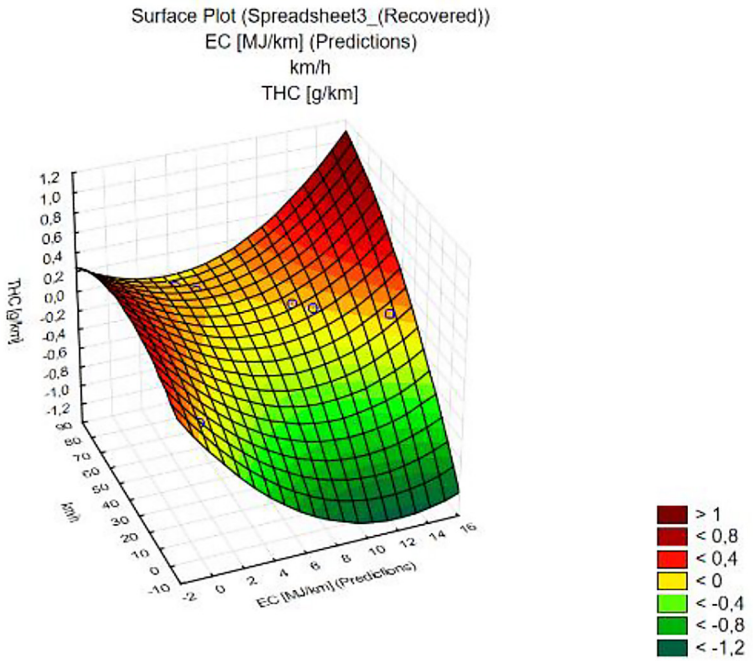

Figure 4. Predicted THC emissions as a function of specific EC and vehicle speed (ML model)

The models made it possible to represent the variability of NMHC emissions depending on the vehicle operating conditions (Figure 6).

NMHC emissions are highest at low speed and high EC, indicating incomplete combustion and inefficient operation of the exhaust gas after-treatment system. A similar increase in emissions is also observed at very high energy consumption, which may result from engine overload. Minimum emission values occur at moderate EC and speeds of 50–70 km/h, confirming the existence of an optimal zone for efficient hydrocarbon combustion.

The use of artificial neural networks, and in particular MLP and SANN models, allows for effective representation of the actual dependencies

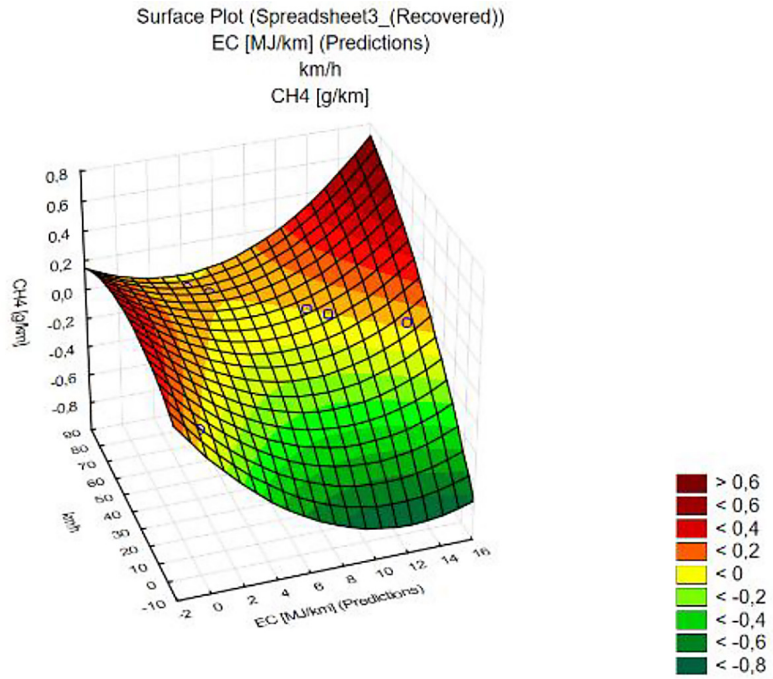

Figure 5. Predicted CH_4 emission depending on the specific EC and vehicle speed

between emissions and energy consumption in the WLTP cycle. These models can be used in:

- real-time prediction of energy consumption (eco-driving),
- digital twin drive systems,
- supporting emission reduction strategies in fleet management systems.

The SANN model proved particularly useful in applications based on time series data, while the MLP model provided the highest overall accuracy. In turn, the SVM can be an alternative in the case of appropriate hyperparameter optimization and the use of feature selection procedures.

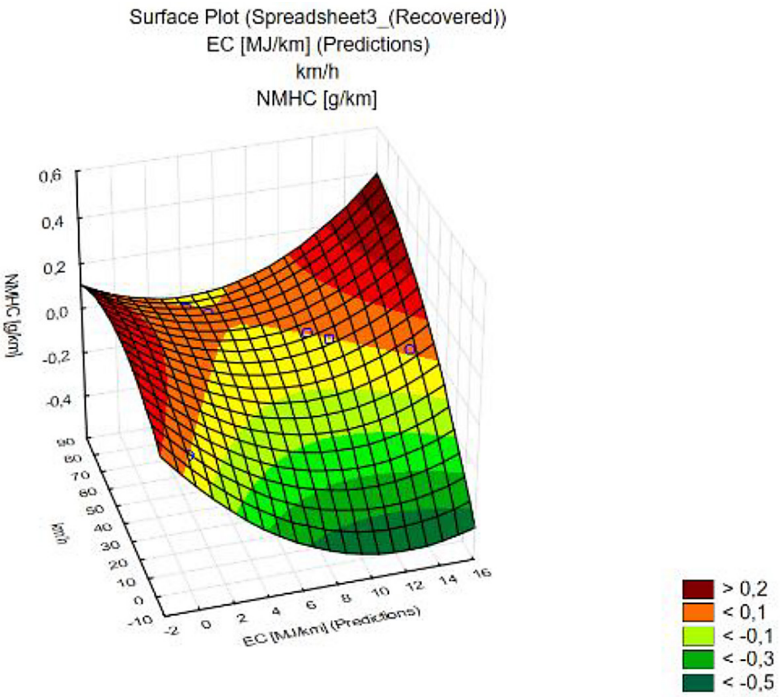

Figure 6. Spatial relationship between specific EC, vehicle speed and predicted NMHC emissions

To facilitate the evaluation of the effectiveness of individual approaches, Table 4 presents a comparison of the most important features and metrics of the three models used: MLP, SANN and SVM.

In summary, integrating experimental data with machine learning methods is a promising avenue for developing accurate forecasting tools for low-emission transportation.

DISCUSSION

The analyses carried out confirm the high effectiveness of artificial intelligence models, in particular MLP and SANN neural networks, in predicting emissions and energy consumption in vehicles with SI engines operating in the WLTP cycle. The MLP model achieved very high prediction accuracy, which is consistent with the results presented by Seo, who reported correlation coefficients above 0.98 for CO and 0.96 for CO_2 [9]. Also in our analysis the relationship between EC and CO_2 emissions was well reproduced, especially in the range of average speeds.

The SANN network has proven to be particularly useful in the analysis of dynamic variables, such as instantaneous speed or THC and CH_4 emissions, whose values are particularly variable during acceleration and braking. According to the literature, models that take into account the temporal nature of data better represent the real emission phenomena in urban conditions. The choice of the fuel system is also crucial for the emission level – non-commercial fuel systems can lead to exceeding the applicable environmental standards, which has been shown in studies on low-power engines.

The SVM model, despite achieving good results in the training set, showed signs of overfitting. This problem is also confirmed by the reports of Zhao et al., who point out the need for careful adjustment of SVM hyperparameters, especially in the analysis of real transportation data [28].

The relationships between emissions and speed and energy consumption were clearly nonlinear, as shown in Figures 2–6. Similar relationships were also identified for diesel vehicles in the study by Seo et al., where the use of a hybrid ANN and vehicle dynamics model enabled accurate prediction of NOx and CO_2 emissions under real operating conditions [29]. For all exhaust components, a tendency for increased emissions was observed under high engine load (high EC) and low driving speed conditions, which is consistent with the results of Fang et al. and Arıkuşu et al., who also observed an increase in emissions during start-up and city driving [30–34].

The obtained results are consistent with the literature reports on the effectiveness of ANN methods in mapping the complex relationships between emissions and fuel consumption. They indicate the need to use a hybrid approach that integrates experimental data and domain knowledge with ML algorithms. Such an approach supports the development of digital twins of vehicles and strategies for optimizing the operation of drive systems. In parallel, the use of AI methods is finding wider application also outside the automotive industry – among others, in the transformation of energy systems, where they support resource optimization, scenario modeling and strategic decision-making. An analogous application of neural networks for energy system forecasting was shown by Orynycz et al. [31], who used ANN and decision trees to predict CHP unit parameters, underlining their potential in hybrid energy systems.

Further research is planned to expand the scope of data to include real road traffic conditions and to implement feature selection and input space optimization methods, which can further increase the accuracy of the models and their application potential.

Table 4. Comparison of machine learning models used to predict energy consumption and pollutant emissions

Model	Data type	Architecture / Parameters	MSE (Train/Test)	R (Train/Test)	SD ratio (Test)	Application
MLP	Table / time	6-10-1, sigmoid + linear, backpropagation	– / – (no numeric values)	high (descriptive)	-	Prediction of EC and emissions as a function of input variables
SANN	Time series	Auto-association, data with delays	– / – (no numeric values)	high (descriptive)	-	Modeling dynamic variables (e.g. speed, emissions)
SVM	Blackboard	C = 7, γ = 0.175, RBF kernel	7.10 / 15.75	0.930 / 0.824	0.610	High accuracy, but susceptible to overfitting

A similar approach based on the analysis of data from actual vehicle operation was used in studies comparing exhaust emissions from engines fueled with diesel oil and LNG, which confirms the validity of extending the data set with such conditions.

Despite the fact that the obtained results indicate the effectiveness of ML models in mapping energy consumption and emissions, it should be emphasized that the dataset is limited (17 observations) [32]. Such a small sample size limits the possibility of generalizing the results and increases the risk of overfitting, especially in the case of models of high complexity (e.g. SVM with RBF). The results should therefore be interpreted as a technical confirmation of the correctness of the proposed methodology. In the next stages, it is planned to expand the database and apply feature selection and cross-validation methods. An important direction of further research may also be the analysis of the energy consumption of individual architectures, in accordance with the approach presented by Tomiło et al.[34]. In future analyses, it is also worth expanding the set of model evaluation metrics to include MAE, RMSE and the coefficient of determination R^2, which will allow for a more accurate assessment of the prediction quality and a comparison of the efficiency of different algorithms.

CONCLUSIONS

The applied machine learning methods – in particular the MLP and SANN models – showed high effectiveness in mapping the relationships between driving speed, energy consumption and pollutant emissions in the WLTP cycle. The MLP model achieved the highest prediction accuracy, while the SANN network was better at analysing dynamic variables such as instantaneous speed or instantaneous emissions.

The SVM model, despite good results on the training set, revealed a tendency to overfitting, which indicates the need for further optimization of hyperparameters or the use of input dimensionality reduction methods. At the same time, its flexibility in parameter selection can be used in selective modeling tasks.

Both statistical analysis and spatial dependence of emissions on driving parameters confirmed the complex, non-linear nature of the EC–emissions relationship. In particular, it was noted that the lowest emissions occurred at moderate speeds and medium engine loads, which is consistent with the physical interpretation of the combustion process in the SI engine.

Integration of experimental data with intelligent predictive algorithms and physicochemical interpretation of combustion processes distinguishes the presented approach from previous studies. The obtained results can be the basis for designing energy management systems in vehicles (eco-driving), digital twins of drive systems and emission reduction strategies in intelligent transport systems.

The proposed methodology can be successfully adapted to the analysis of emissions from vehicles powered by alternative fuels, in other test cycles (e.g. RDE) or in real road traffic conditions.

Further research is planned to expand data sets, introduce real operating scenarios, and compare ANN models with alternative ML algorithms. The research results are a significant step towards intelligent modeling of vehicle emissions and energy in the context of sustainable transport development.

In further work it will be necessary to increase the number of observations and use advanced validation techniques, which will enable a more reliable assessment of the predictive effectiveness of the proposed models.

Acknowledgments

The research was carried out with the financial support obtained from the research subsidy of the Faculty of Engineering Management (WIZ) of Bialystok University of Technology, grant No. WZ/WIZ-INZ/2/2025 (Olga Orynycz).

REFERENCES

1. Lejri D, Can A, Schiper N, Leclercq L. Accounting for traffic speed dynamics when calculating COPERT and PHEM pollutant emissions at the urban scale. Transportation Research Part D: Transport and Environment 2018;63:588–603. https://doi.org/10.1016/j.trd.2018.06.023
2. Samaras C, Tsokolis D, Toffolo S, Magra G, Ntziachristos L, Samaras Z. Enhancing average speed emission models to account for congestion impacts in traffic network link-based simulations. Transportation Research Part D: Transport and Environment 2019;75:197–210. https://doi.org/10.1016/j.trd.2019.08.029

3. Kulisz M, Kujawska J, Cioch M, Cel W. Modeling energy access challenges in Europe: a neural network approach to predicting household heating inadequacy using macro-energy indicators. Energies 2024;17:6104. https://doi.org/10.3390/en17236104
4. Surblys V, Kozłowski E, Matijošius J, Gołda P, Laskowska A, Kilikevičius A. Accelerometer-based pavement classification for vehicle dynamics analysis using neural networks. Applied Sciences 2024;14:10027. https://doi.org/10.3390/app142110027
5. Haykin S. Neural networks: a comprehensive foundation. Prentice Hall PTR. 1994.
6. Goodfellow I. Deep learning. n.d.
7. Zimakowska-Laskowska M, Laskowski P, Wojs MK, Orliński P. Prediction of pollutant emissions in various cases in road transport. Applied Sciences 2022;12:11975. https://doi.org/10.3390/app122311975
8. Zimakowska-Laskowska M, Kozłowski E, Laskowski P, Wiśniowski P, Świderski A, Orynycz O. Vehicle exhaust emissions in the light of modern research tools: synergy of chassis dynamometers and computational models. Combustion Engines 2025;200:145–54. https://doi.org/10.19206/CE-201224
9. Seo J, Lim Y, Han J, Park S. Machine learning-based estimation of gaseous and particulate emissions using internally observable vehicle operating parameters. Urban Climate 2023;52:101734. https://doi.org/10.1016/j.uclim.2023.101734
10. Fu J, Yang R, Li X, Sun X, Li Y, Liu Z, et al. Application of artificial neural network to forecast engine performance and emissions of a spark ignition engine. Applied Thermal Engineering 2022;201:117749. https://doi.org/10.1016/j.applthermaleng.2021.117749
11. Hashemi N, Clark NN. Artificial neural network as a predictive tool for emissions from heavy-duty diesel vehicles in Southern California. International Journal of Engine Research 2007;8:321–36. https://doi.org/10.1243/14680874JER00807
12. Valeika G, Matijošius J, Orynycz O, Rimkus A, Kilikevičius A, Tucki K. Compression ignition internal combustion engine's energy parameter research using variable (HVO) biodiesel and biobutanol fuel blends. Energies 2024;17:262. https://doi.org/10.3390/en17010262
13. Kong X, Chen Z, Liu W, Ning K, Zhang L, Muhammad Marier S, et al. Deep learning for time series forecasting: a survey. Int J Mach Learn & Cyber 2025. https://doi.org/10.1007/s13042-025-02560-w
14. Shin S, Lee Y, Kim M, Park J, Lee S, Min K. Deep neural network model with Bayesian hyperparameter optimization for prediction of NO x at transient conditions in a diesel engine. Engineering Applications of Artificial Intelligence 2020;94:103761. https://doi.org/10.1016/j.engappai.2020.103761
15. Tucki K, Orynycz OA, Świć A, Wasiak A, Mruk R, Gola A. Analysis of the possibility of using neural networks to monitor the technical efficiency of diesel engines during operation. Adv Sci Technol Res J 2023;17:1–15. https://doi.org/10.12913/22998624/172003
16. Sun P, Cheng X, Yang S, Ruan E, Zhao H, Liu Y, et al. Artificial neural network models for forecasting the combustion and emission characteristics of ethanol/gasoline DFSI engines with combined injection strategy. Case Studies in Thermal Engineering 2024;54:104007. https://doi.org/10.1016/j.csite.2024.104007
17. Sathishkumar S, Ibrahim MM. Evaluation of machine learning algorithms on hydrogen boosted homogeneous charge compression ignition engine operation for performance and emission prediction. Process Safety and Environmental Protection 2025;195:106756. https://doi.org/10.1016/j.psep.2025.01.010
18. Tasdemir S, Saritas I, Ciniviz M, Allahverdi N. Artificial neural network and fuzzy expert system comparison for prediction of performance and emission parameters on a gasoline engine. Expert Systems with Applications 2011:S0957417411007251. https://doi.org/10.1016/j.eswa.2011.04.198
19. Yang R, Yan Y, Sun X, Wang Q, Zhang Y, Fu J, et al. An artificial neural network model to predict efficiency and emissions of a gasoline engine. Processes 2022;10:204. https://doi.org/10.3390/pr10020204
20. Azeez OS, Pradhan B, Shafri HZM, Shukla N, Lee C-W, Rizeei HM. Modeling of CO emissions from traffic vehicles using artificial neural networks. Applied Sciences 2019;9:313. https://doi.org/10.3390/app9020313
21. Alazemi F, Alazmi A, Alrumaidhi M, Molden N. Predicting fuel consumption and emissions using GPS-Based machine learning models for gasoline and diesel vehicles. Sustainability 2025;17:2395. https://doi.org/10.3390/su17062395
22. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97. https://doi.org/10.1007/BF00994018
23. Dini P, Basso G, Saponara S, Romano C. Real-time monitoring and ageing detection algorithm design with application on SiC-based automotive power drive system. IET Power Electronics 2024;17:690–710. https://doi.org/10.1049/pel2.12679
24. Tlelo-Cuautle E, Rangel-Magdaleno JDJ, De La Fraga LG. Artificial Neural Networks for Time Series Prediction. Engineering Applications of FPGAs, Cham: Springer International Publishing; 2016, p. 117–50. https://doi.org/10.1007/978-3-319-34115-6_5

25. Gierz Ł, Al-Sammarraie MAJ, Özbek O, Markowski P. The use of image analysis to study the effect of moisture content on the physical properties of grains. Sci Rep 2024;14. https://doi.org/10.1038/s41598-024-60852-7
26. Al-Sammarraie MAJ, Gierz ŁA, Özbek O, Kırılmaz H. Power predicting for power take-off shaft of a disc maize silage harvester using machine learning. Adv Sci Technol Res J 2024;18:1–9. https://doi.org/10.12913/22998624/188666
27. Hussain I, Ching KB, Uttraphan C, Tay KG, Noor A. Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: A comparative study. Sci Rep 2025;15:16124. https://doi.org/10.1038/s41598-025-94946-7
28. Zhao D, Li H, Hou J, Gong P, Zhong Y, He W, et al. A review of the data-driven prediction method of vehicle fuel consumption. Energies 2023;16:5258. https://doi.org/10.3390/en16145258
29. Seo J, Yun B, Park J, Park J, Shin M, Park S. Prediction of instantaneous real-world emissions from diesel light-duty vehicles based on an integrated artificial neural network and vehicle dynamics model. Science of The Total Environment 2021;786:147359. https://doi.org/10.1016/j.scitotenv.2021.147359
30. Fang L, He B. A deep learning framework using multi-feature fusion recurrent neural networks for energy consumption forecasting. Applied Energy 2023;348:121563. https://doi.org/10.1016/j.apenergy.2023.121563
31. Orynycz O, Zimakowska-Laskowska M, Kulesza E. CO2 Emission and Energy Consumption Estimates in the COPERT Model—Conclusions from Chassis Dynamometer Tests and SANN Artificial Neural Network Models and Their Meaning for Transport Management. Energies 2025;18:3457. https://doi.org/10.3390/en18133457
32. Qiao Q, Eskandari H, Saadatmand H, Sahraei MA. An interpretable multi-stage forecasting framework for energy consumption and CO_2 emissions for the transportation sector. Energy 2024;286:129499. https://doi.org/10.1016/j.energy.2023.129499
33. Tomiło P. Classification of the condition of pavement with the use of machine learning methods. Transport and Telecommunication Journal 2023;24:158–66. https://doi.org/10.2478/ttj-2023-0014