

## Article

# Sensor-Based Machine Learning Approach for Indoor Air Quality Monitoring in an Automobile Manufacturing

Yose Wandy <sup>1,\*</sup>, Marcus Vogt <sup>2</sup>, Rushit Kansara <sup>2</sup>, Clemens Felsmann <sup>1</sup> and Christoph Herrmann <sup>2</sup>

<sup>1</sup> Chair of Building Energy Systems and Heat Supply, Institute of Energy Technology, Technische Universität Dresden, Helmholtzstr. 14, 01069 Dresden, Germany; clemens.felsmann@tu-dresden.de

<sup>2</sup> Chair of Sustainable Manufacturing and Life Cycle Engineering, Institute of Machine Tools and Production Technology, Technische Universität Braunschweig, Langer Kamp 19b, 38106 Braunschweig, Germany; marcus.vogt@tu-braunschweig.de (M.V.); rushit2011@gmail.com (R.K.); c.herrmann@tu-braunschweig.de (C.H.)

\* Correspondence: ywandy@ymail.com

**Abstract:** The alternative control concept using emission from the machine has the potential to reduce energy consumption in HVAC systems. This paper reports on a study of alternative inputs for a control system of HVAC using machine learning algorithms, based on data that are gathered in a welding area of an automotive factory. A data set of CO<sub>2</sub>, fine dust, temperatures and air velocity was logged using continuous and gravimetric measurements during two typical production weeks. The HVAC system was reduced gradually each day to trigger fluctuations of emission. The data were used to train and test various machine learning models using different statistical indices, consequently to choose a best fit model. Different models were tested and the Long Short-Term Memory model showed the best result, with 0.821 discrepancy on R<sup>2</sup>. The gravimetric samples proved that the reduction of air exchange rate does not correlate to escalation of fine dust linearly, which means one cannot rely on just gravimetric samples for HVAC system optimization. Furthermore, by using machine learning algorithms, this study shows that by using commonly available low cost sensors in a production hall, it is possible to correlate fine dust data cost effectively and reduce electricity consumption of the HVAC.

**Keywords:** machine learning; HVAC; control system; body shop; automotive industry



**Citation:** Wandy, Y.; Vogt, M.; Kansara, R.; Felsmann, C.; Herrmann, C. Sensor-Based Machine Learning Approach for Indoor Air Quality Monitoring in an Automobile Manufacturing. *Energies* **2021**, *14*, 7271. <https://doi.org/10.3390/en14217271>

Academic Editors: Wei-Hsin Chen, Núria Agell, Zhiyong Liu, Ying-Yi Hong and Boris Igor Palella

Received: 29 July 2021

Accepted: 20 October 2021

Published: 3 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The digitalization of manufacturing is being driven forward by the decreasing costs of information and communication technology. Cyber-physical systems (CPS), or in the manufacturing context Cyber-Physical Production Systems (CPPS), are an important technological element in the realization of the 4th industrial revolution [1]. CPPS and their applicability in industrial environments are of increasing interest in current research and industry. A CPPS is clustered into subsystems: the physical world and the cyber world. These subsystems interact with each other through data acquisition and decision-making support and control functionalities, respectively [2].

In 2018, the World's energy consumption has increased by almost 40% compared to the consumption in 1990 [3]. The industrial share of total final consumption (TFC) stayed at the ~30% [4]. Although the absolute consumption increased to ~57% compared to the consumption of industrial sector in 1990 [5]. In times of increasing energy demand and decreasing energy resources, a further optimization and revaluation in industrial sector is imperative.

The Heating Ventilating and Air-Conditioning (HVAC) system is responsible for a large part of the energy consumption in industry. According to internal Volkswagen data [6], it accounts on average for about 40% of the total energy consumption in industrial buildings. Automobile production can be classified into press shop, assembly, paint shop

and body shop. Body shop is responsible for almost one-quarter of the total energy consumption, preceded only by the paint shop [6]. Within the case study of a body shop, the HVAC system has the highest energy demand, succeeding the other infrastructure consumption: the hall lighting and workplace lighting. In the body shop, the HVAC consumption exceeds 60% of the total energy demand, succeeding the consumption from the production itself [6].

The main function of an HVAC system is to satisfy and maintain the requirements with respect to air and building environment in general, for example, thermal comfort, air pollution control and the hygienic aspect, which is achieved by conditioning outdoor air to the desired levels in occupied buildings or for product processing and transporting the air in the room through an exhaust duct. It controls and maintains temperature, humidity, air movement, air cleanliness and pressure differential within a defined space. Wang [7] classified HVAC systems into comfort HVAC systems and process HVAC systems depending on the application. Comfort HVAC systems provide occupants with a comfortable and healthy indoor climate and process HVAC systems provide and maintain the necessary air conditioning for production, storage or other defined processes [7,8].

In order to guarantee that the indoor climate in the body shop is up to the user requirements, the production's emissions need to be transported out of the conditioned space using the HVAC system, this is done using the technical exhaust air vent. One of the main emissions in a body shop is fine dust, which is a byproduct of the welding process. In the geographical location of this study, Wrzesnia, Poland, to be precise, an occupational exposure limit is regulated under TRGS 900 [9]. The limit values are  $1.25 \text{ mg/m}^3$  for PM<sub>2.5</sub> and  $10 \text{ mg/m}^3$  for PM<sub>10</sub>; this is then taken up again in Chapter 2 and elaborated on in more detail.

Fine dust can be measured using specific measuring technology, specifically using an optical sensor that measures and monitors fine dust concentrations with preconditioned air. Within the study, the fine dust is measured by this optical scattered light measurement [10]. However, this method is highly cost-intensive [11].

Digital technologies, such as the Internet of Things (IoT), offer a wide number of applications in various sectors [12]. One of the examples is to use cost-effective smart sensors, which enable users to measure ambient conditions and upload them to the cloud [13]. This enables users to experiment or use a more modern control concept, for example, machine learning algorithms. Li [14] presented a hybrid model based on long short-term memory (LSTM) network and attention mechanism (LSTM-attention) applied to the prediction of Total Suspended Particulate Matter (TSP) concentration. Li [14] showed that the model approach can be applied to the prediction of dust concentrations in open-pit mines, thus helping in decision support on when to carry out dust suppression work.

The main scope of this paper is to study different model approaches using data measured in an industrial welding area of an automobile body shop.

The paper starts with a literature review concerning fine dust or particulate matter and the national regulations and technical rules concerning Particulate Matter (PM) and industrial welding. It then explains the methodology and approach used in this paper. This is followed by the factory data collection setup, data analysis and data correlations. Subsequently, the created data models or algorithms are presented with the results and the statistical comparisons. Finally, the conclusion and future scope or applicability of the paper are outlined. The project shows the possibilities of using an alternative data model correlation from more common and less expensive parameters, such as air velocity, air temperature, relative humidity, the CO<sub>2</sub> content around the perimeter and the electric current of the welder. By using machine learning algorithms, we show that by using commonly available sensors in a production hall, it is possible to correlate fine dust data cost-effectively.

## 2. Literature Review

### 2.1. Particulate Matter (PM)

Fine dust or Particulate matter (PM) is the term used to describe particles in the air that do not sink to the ground immediately, but remain in the atmosphere for a certain period of time. Particulate matter can come either from natural sources or from human activities. Natural contributions include emissions from volcanoes, oceans, soil erosion, forest and bush fires, as well as dead and eroded (abraded) organic remains, pollen, spores, bacteria and viruses. Sources from human activities include transport, power plants, district heating plants, waste incineration plants, private and commercial heating systems and certain industrial processes such as metal production, especially steel production [15,16].

Airborne particles range in size from a few nanometers (nm) to about 100 micrometers ( $\mu\text{m}$ ). This depends on the process and material used in the working environment. Madanchi [17] presented an experimental approach to evaluate the influence of process parameters on the formation of emissions and particle size. The work showed that the process parameter (cutting speed and cooling lubricant volume flow) has linear impact not only towards the material removal process, but also the emissions and particle size.

The size of the dust particles, characterized by “aerodynamic diameter” as a model, is the decisive property for their classification [15,18]:

- The term ‘Total Suspended Particulate Matter’ (TSP) covers all suspended particles with an aerodynamic diameter smaller than  $\sim 60$  micrometers ( $\mu\text{m}$ ),
- PM10 refers to all particles with aerodynamic diameter less than 10 micrometers.
- PM2.5 refers to particles with an aerodynamic diameter of less than 2.5 micrometers.
- The smallest categorized particles are called ultra fine particles. Their aerodynamic diameters are less than 0.1 micrometers.

The nose, mouth and throat can filter back particles larger than 10 micrometers. A small percentage of particles between 10 and 2.5  $\mu\text{m}$  (PM10-2.5) are able to reach the bronchiole and alveoli. Particles that have a diameter smaller than 2.5  $\mu\text{m}$  (PM2.5) have a higher probability to reach and stay in the alveoli. These particles can trigger inflammatory processes in the lung tissue [19]. In the alveoli, respiration and blood circulation are functionally and anatomically very closely connected [15]. This is why abnormalities in one system, such as inflammations in the respiratory tract, can also additionally affect other systems, for example, the heart or circulatory system [15].

### 2.2. National Regulations and Technical Rules concerning Particulate Matter (PM) and Industrial Welding

The main focus of welding processes lies on materials (filler and base material). However, note that neither base materials nor filler materials (for example, welding materials) are considered “hazardous substances”. “Hazardous substances” are only released as a result of a welding process, which consist primarily of particulate and/or gaseous hazardous substances. These hazardous substances are classified within the meaning of the Regulation on Hazardous Substances (GefStoffV). An occupational exposure limit was included in TRGS 900 [9]. TRGS 900 shows a list of the quantities’ threshold for substances with specific toxic properties, for example, for dusts containing heavy metals or for wood dusts and flour dusts. If the substance in use is not listed, then a general threshold of fine dust is applicable. These thresholds are compulsory in Germany from beginning of 2019, after a transitional period of five years (until 2018) was established. However, welding fumes are not dusts in the sense of the definition, as dusts are produced during mechanical processes, for example grinding. Respectively, welding fumes are produced by thermal processes at very high temperatures, for example, arc welding. Welding fumes can therefore not be directly equated with dust. It consists of mainly from particles that are alveolar. In comparison to fine dust, fumes are smaller in size, approximately between 0.2 to 1  $\mu\text{m}$  [20]. This means that the welding fume particles and in unfavorable workplace situations to impair the ability of the person exposed to the respiratory tract or damage it. Furthermore, other negative health effects on the human organism are also to be considered [21]. A direct

specific threshold for welding fumes has not yet been defined. The metal compound that can be found in welding fumes has a dust density higher than  $0.0025 \text{ mg/m}^3$ , approximately between  $0.005 \text{ mg/m}^3$  and  $0.007 \text{ mg/m}^3$  [22]. In reality, this is the reason for the risk assessment: the dust limit values are measured and converted mathematically to get the density of the metals from the fumes [22].

Basically, the dust limit value for alveolar dust is ( $<2.5 \mu\text{m}$ ) of  $1.25 \text{ mg/m}^3$  [9]. DIN EN 689 [23] defined the limit further, including the area analysis, control measurements, protocol and processing of measurement data. In the automotive industry, companies such as Volkswagen AG have to establish their safety measures and define their time interval between control measures, based on their measurements data and national regulations. According to DIN EN 689 [23], the time interval between measurements should be determined taking into account the following factors:

- process cycles taking into account normal working conditions;
- effects of failure of control devices;
- proximity to the limit value;
- the effectiveness of process controls;
- time span until the condition is under control again;
- the variation over time of the measurement results.

Based on these factors and the upper limit defined by the national regulations, Volkswagen AG detailed their guidelines concerning the threshold and control measurements in welding area for alveolar dust. Using gravimetric analysis, the expositions in welding areas are measured and controlled.

### 3. Methodology and Approach

As methodological framework a Cyber-Physical Production Systems (CPPS) is used; Figure 1 shows an adapted CPPS framework for this application.

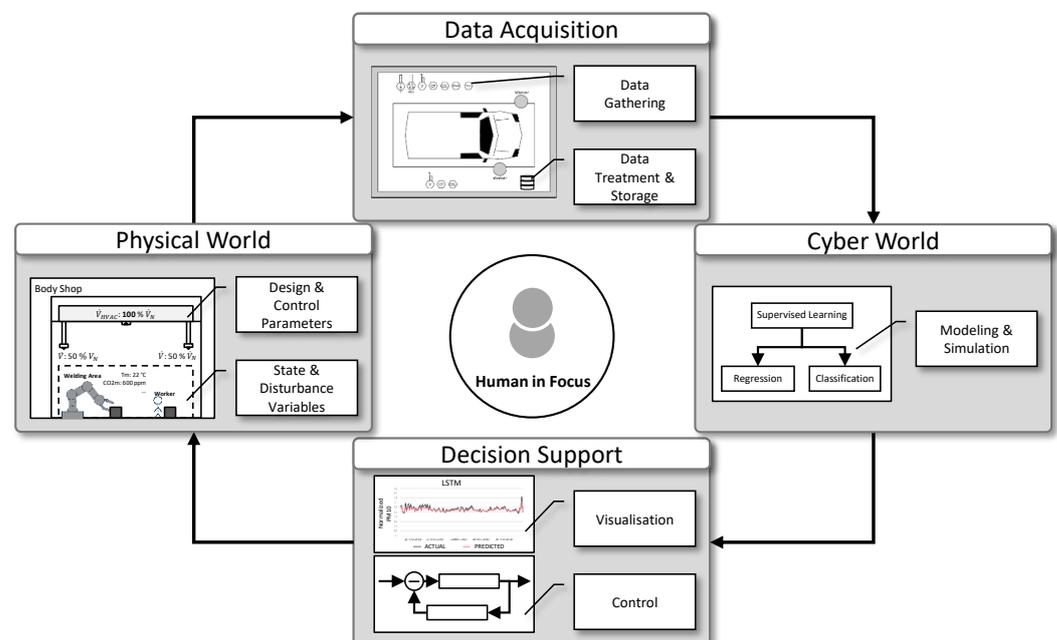


Figure 1. Schematic of CPPS framework for the use case based on the work in [2].

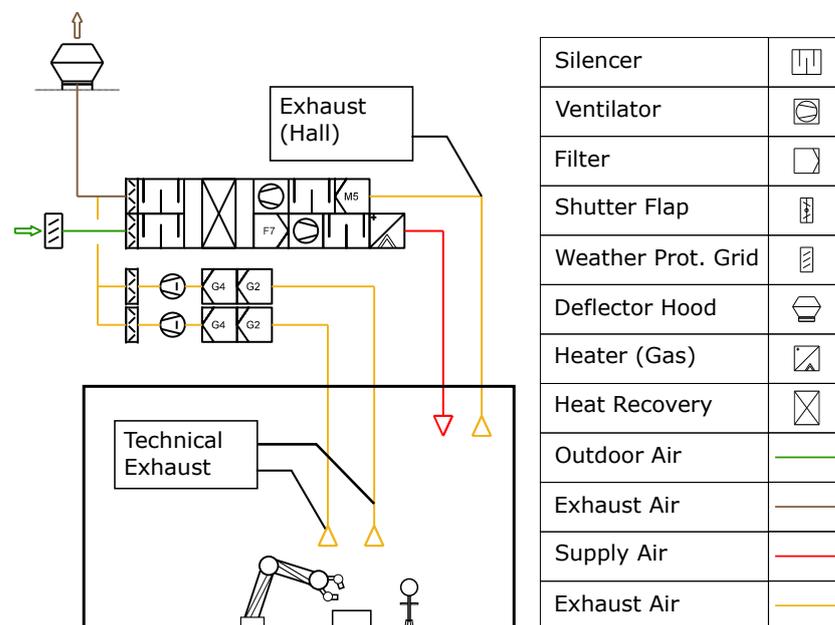
The approach to integrate a CPPS in the body shop is driven by the goals of improving energy efficiency while maintaining air quality in the production environment and being cost efficient. The CPPS system provides a decision support by calculating the fine dust concentration in the production environment. This enables the HVAC operator in the physical world to use conservative sensors, such as temperature, relative humidity and

CO<sub>2</sub> to operate the HVAC system. Further, the CPPS enables a model-based control in the future.

With this cyber-physical HVAC system, energy efficiency, air quality and thermal conditions can be improved in the production environment by providing decision support, greater transparency and enabling model-predictive control in real-time. With the proposed approach, one can measure fine dust in the air without using expensive fine dust sensor in each welding area in a body shop. This can lead to a high cost savings when monitoring a large body shop hall in a factory. The low-cost IoT sensor technology combined with machine learning enables large-scale deployment with moderate costs compared to the high-precision fine dust sensors on factory scale.

### 3.1. Setup for Collecting Data in the Factory Hall

In order to find the correlation between the HVAC-System, air quality and production activity, it is necessary to collect and analyze the variables on site. An experiment within the scope of a collaboration project between TU Braunschweig and Volkswagen AG is thus commenced. To investigate seasonal influences, experiments were conducted in winter and summer at five consecutively working days, to be exact from 6th to 8th of August 2018 and from 26th to 30th of November 2018. This allowed us to measure the experiment under 5 different air exchange rates, which were set 8 h before the start of each measurement. For the first three days, the reductions were concentrated on the main exhaust air vents, the openings of which can be found directly under the ceiling of the production hall, this is detailed Table 1. Moreover, for the last two days, we set to measure the effect on the reductions on the technical exhaust vents, which is directly responsible for exhausting the welding emission. Figure 2 shows the difference on the positions of the two exhaust vents and the HVAC system used in the factory hall.



**Figure 2.** Schematic representation of the HVAC system and the difference between the two exhausts.

The reduction limit was set to 50% because the technical exhaust vents are the main way to exhaust the byproducts of the welding. Table 1 lists the experimental parameters for the work days with the expected energy saving potential in compare to the reference day (monday), which is based on the assumption of constant fan efficiency. A change in volume flow rate affects the fan power requirement as follows [8]:

$$P_2 = P_1 \left( \frac{V_2}{V_1} \right)^3$$

for example, 10% less volume rate will lower power demand by 33%. The supply air fan has a nominal power of 40 kW, while the exhaust air fan and the technical exhaust fan operate at nominal power of 30 kW and 60 kW.

**Table 1.** Parameters for the experiments in winter and summer with the theoretical electricity savings.

Day	Supply Air	Energy Saving	Exhaust Air	Energy Saving	Techn. Exhaust	Energy Saving
Monday	75.000 m <sup>3</sup> /h	-	22.137 m <sup>3</sup> /h	-	24.000 m <sup>3</sup> /h	-
Tuesday	-13%	34%	-46%	84%	-0%	0%
Wednesday	-30%	66%	-100%	100%	-0%	0%
Thursday	-16%	41%	-0%	0%	-30%	66%
Friday	-17%	43%	-0%	0%	-50%	88%

The samples were taken during welding corrections; welding of floor elements (galvanized steel) with the welding torch SKS Power Feeder PF5 (MAG method); coated with steel wire with copper with a diameter of 1.0 mm; Argon; and CO<sub>2</sub> inert gas welding, sanding and cleaning of welding seams. An oscillating time of at least 6 h was taken into account, so that the air conditions in the welding area settle before the measurements were commenced.

### 3.2. Data Acquisition and Analysis

In order to ensure that the regulatory thresholds are well maintained, certified gravimetric measurements from the work safety department accompanied the experiment. This gravimetric measurements were executed according to PN EN 689:2018, a Polish version of EN 689 [24]. Tables 2 and 3 show the result of the gravimetric measurements respectively in summer time between 6 and 8 August 2018 and winter time from 26 to 30 November 2018, measured over the time period of 8 h. The gravimetric measurement devices were hung directly on the workers shown in Figure 3. The results of the gravimetric measurements are shown in Tables 2 and 3.

**Table 2.** Results from the gravimetric measurements in summer.

	06.08.2018	07.08.2018	08.08.2018	09.08.2018	10.08.2018
PM2.5 (mg/m <sup>3</sup> )	0.29	3.24	0.89	0.36	0.31
PM10 (mg/m <sup>3</sup> )	<0.11	1.42	1.52	0.42	0.2

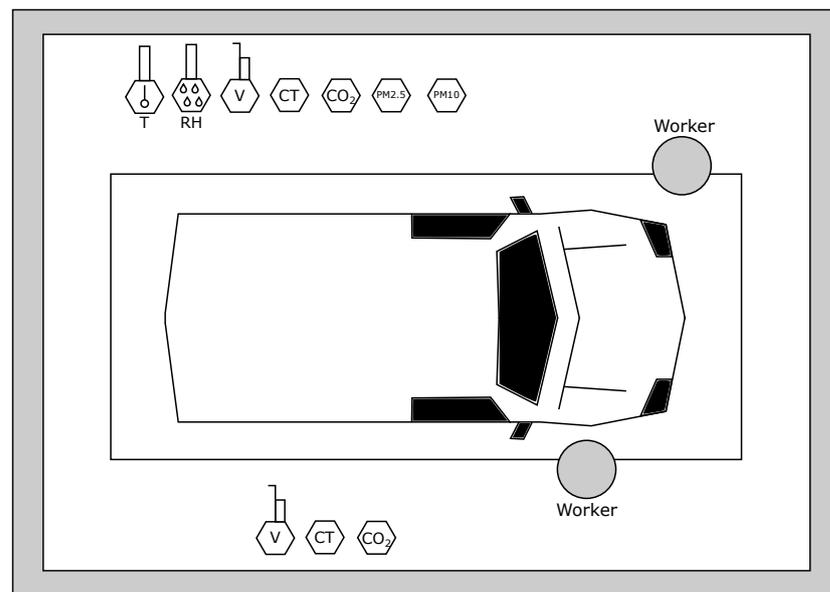
The exposition limits (PM2.5 = 1.25 mg/m<sup>3</sup> and PM10 = 10 mg/m<sup>3</sup>) were held under the regulatory threshold during the experiments, except for the PM2.5 on Tuesday 7 August 2018. This outlier, however, was not backed by the PM10 result on the same day, which stayed far below the limit of 10 mg/m<sup>3</sup>. The measurements with the same air exchange rate settings were also repeated in winter and shown in the Table 3. The measurement results are more dependent on workers behavior, movements and routines, rather than the reduction in air exchange rate shown in Table 1, which might explain the spike and the nonlinear results.

**Table 3.** Results from the gravimetric measurements in winter.

	26.11.2018	27.11.2018	28.11.2018	29.11.2018	30.11.2018
PM2.5 (mg/m <sup>3</sup> )	0.43	0.26	<0.12	0.39	<0.12
PM10 (mg/m <sup>3</sup> )	0.58	0.38	0.38	0.58	0.22

### 3.3. Correlation Study between Fine Dust and Other Parameters

In order to be able to design a system which can predict the amount of values of PM10 and PM2.5 with the combination of other sensors, reference data sets are measured and collected.



**Figure 3.** Schematic representation of the experimental measurements (V = Air Velocity, CT = Electric Current used for welding, T = Air Temperature, RH = Relative Humidity, CO<sub>2</sub> = Carbon Dioxide).

Figure 3 shows a representative schematic of the measurements setup in the hand welding area on-site. The places chosen for the sensor installations represent the spots where workers operate manually using a hand welder during the productions, which means the resulting air velocity, Temperature, Relative Humidity and CO<sub>2</sub> depend greatly on workers actions and activities. The fine dust PM2.5 and PM 10 are measured using optical sensors as well providing a dataset for the purpose of validating the designed algorithm. The sensor used was fine dust FDS 15 from Dr. Födisch Umweltmesstechnik AG which has accuracy of  $\pm 5 \mu\text{g}/\text{m}^2$  and a measuring range from  $2 \mu\text{g}/\text{m}^2$  to  $3.000 \mu\text{g}/\text{m}^2$  [10].

Table 4 shows the results of the measurements data in winter time on 8-h average, the data of PM2.5 and PM10 were measured using optical sensors FDS 15. With the proper use of the data from these sensors combined with an algorithm, the amount of PM10 and PM2.5 in the air can be predicted to a certain extent. Which variables are to be selected for input can be decided based on the correlation analysis. The best combination of the inputs, that gives the lowest error in the prediction, can be chosen as input variables for the model. This selection is known as feature selection [25].

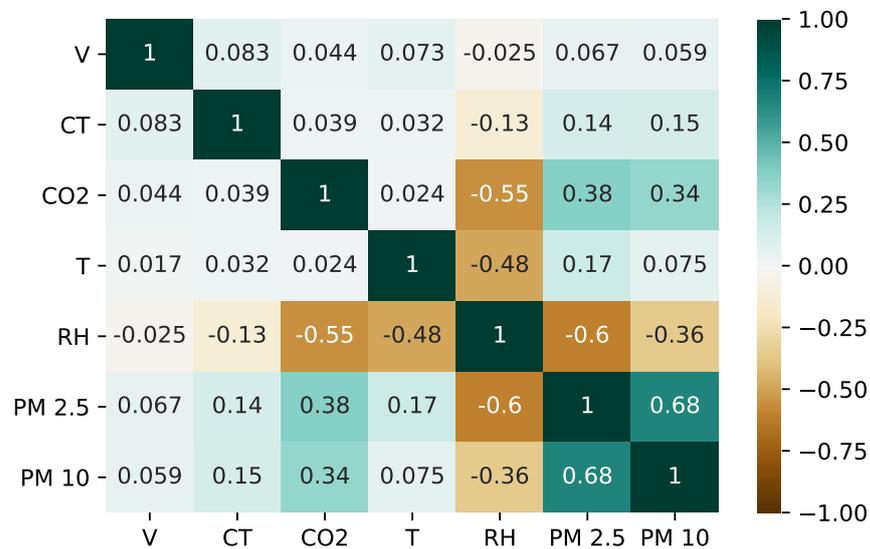
From the correlation analysis, the dependency of PM10 and PM2.5 on other available variables is noticeable. Figures 4 and 5 show the calculated linear coefficients for each variable using Pearson and Spearman's Rho Correlation. The correlation between and input variables and target variable provides the basis for the feature extraction. Strong correlation means that the changes in chosen features (variables) causes higher changes in the target variable. Pearson coefficient suggests how two variables are correlated linearly. On the other hand, Spearman's Rho coefficient shows the nonlinear monotonic relation between two variables. The formula for the Spearman rank correlation coefficient is [26]

$$r_s = 1 - \left( \frac{6 \sum d_i^2}{n(n^2 - 1)} \right),$$

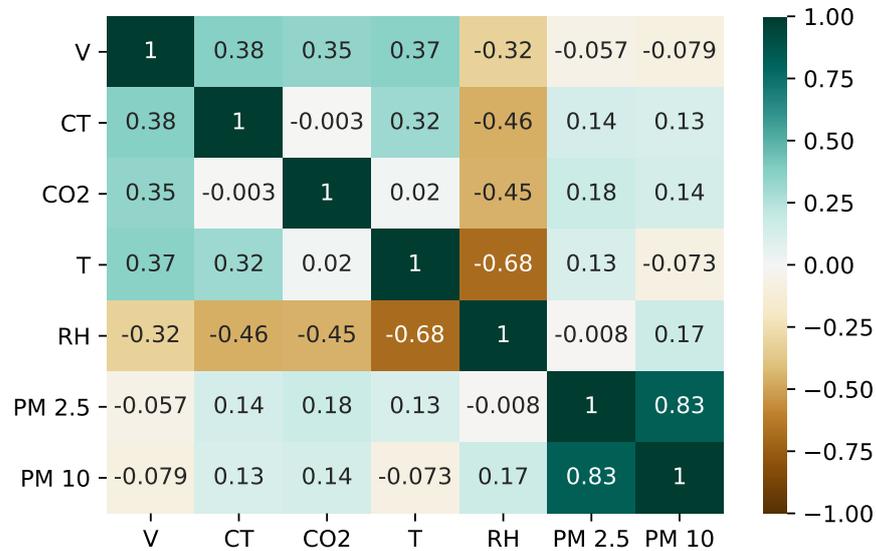
where  $d_i = R(X_i) - R(Y_i)$  is the difference between two ranks of each observation and  $n$  is the number of observations.

**Table 4.** 8-h average values of the field data measurements.

Time	Air Vel. [m/s]	Current [A]	CO <sub>2</sub> [ppm]	RH [%]	T [°C]	PM2.5 [mg/m <sup>3</sup> ]	PM10 [mg/m <sup>3</sup> ]
26 Nov 2018 08:00	0.15	2.07	855.2	34.9	21.8	0.063	0.084
26 Nov 2018 16:00	0.14	1.27	877.7	31.5	21.3	0.062	0.088
27 Nov 2018 00:00	0.13	1.30	966.5	29.8	21.3	0.061	0.091
27 Nov 2018 08:00	0.13	1.15	876.4	30.7	21.5	0.061	0.081
27 Nov 2018 16:00	0.13	1.24	847.1	30.3	21.6	0.061	0.083
28 Nov 2018 00:00	0.13	1.09	869.3	31.5	21.6	0.061	0.086
28 Nov 2018 08:00	0.13	1.25	907.4	31.1	21.6	0.064	0.079
28 Nov 2018 16:00	0.15	1.72	911.9	29.5	21.6	0.080	0.086
29 Nov 2018 00:00	0.18	1.31	999.2	26.9	21.7	0.077	0.083
29 Nov 2018 08:00	0.18	1.17	943.3	25.0	21.8	0.076	0.082
29 Nov 2018 16:00	0.21	1.57	944.2	24.6	21.8	0.076	0.082
30 Nov 2018 00:00	0.20	1.68	957.9	25.6	21.8	0.075	0.080
30 Nov 2018 08:00	0.24	3.31	890.7	26.2	21.9	0.077	0.081



**Figure 4.** Correlation analysis using Pearson Correlation.



**Figure 5.** Correlation analysis using Spearman's Rho Correlation.

Table 5 shows exemplary  $d_i$  as the difference between ranks of each observation of PM2.5 and relative humidity (RH), where  $n$  defined the total number of observations, which means total number of samples, in the case of this study approximately 350,000.

**Table 5.** Spearman's Rho Correlation Formula on PM2.5 and Relative Humidity (RH).

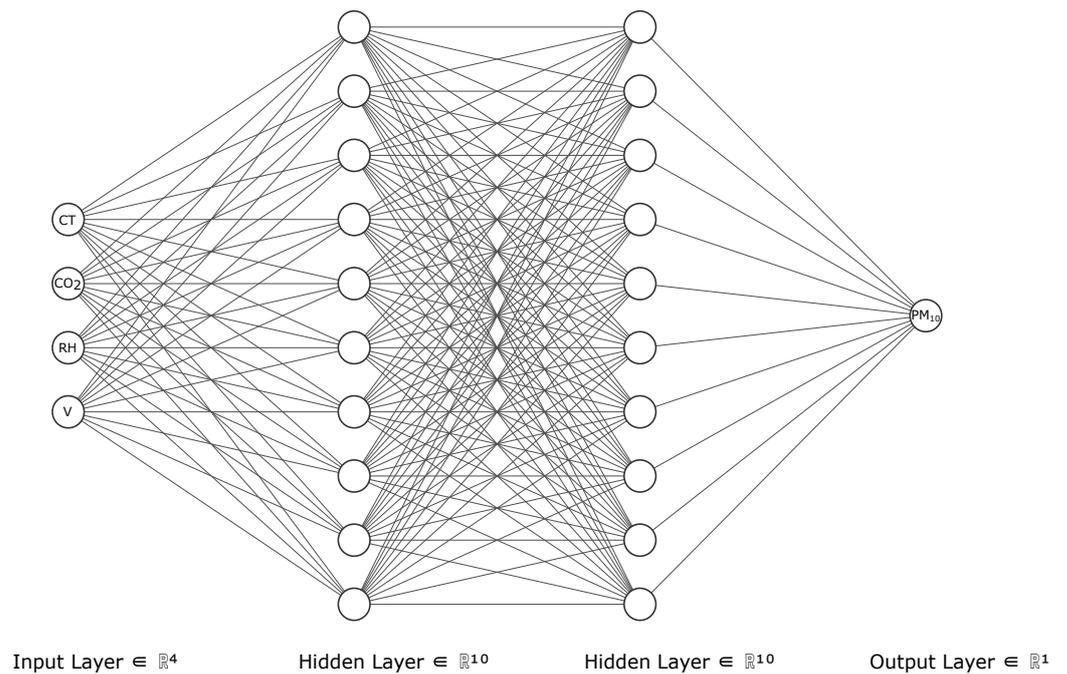
Observations	PM2.5 Rank	RH Rank	d
1	10	7	3
2	124	80	44
...	...	...	...

Each variable has correlation a coefficient of 1 to itself. Because both substances are categorized as fine dust, PM10 and PM2.5 have a strong positive linear correlation to each other. Both substances have negative intermediate correlations with relative humidity (RH) and positive intermediate correlations with CO<sub>2</sub>. The electricity current (CT) shows a weak positive correlation to both PM10 and PM2.5. These variables were chosen, as those are the potential quantities which might affect the amount of PM2.5 and PM10 in the welding area.

### 3.4. Data Model

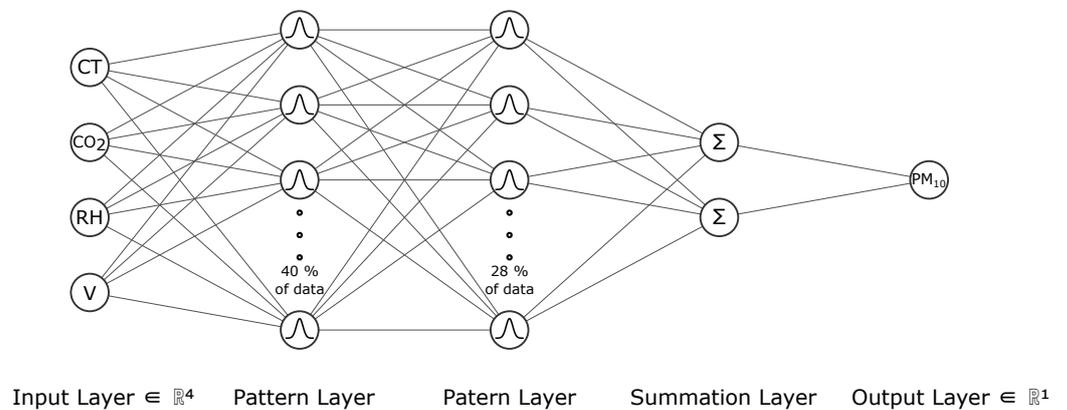
Spearman's rho coefficient test is used to show the monotonic behavior of the variables to each other. PM10 and PM2.5 show a strong increasing monotonic relation with each other and also have sufficient monotonic relation with the current (CT), CO<sub>2</sub> and relative humidity (RH). This result also shows that PM10 and PM2.5 have decreasing monotonic behavior with Air Velocity (V). The combination of Pearson and Spearman's coefficients show that the Current (CT), CO<sub>2</sub>, Relative Humidity (RH) and Air Velocity (V) are more suitable to serve as input variables for the model. The aim is to exclude PM2.5 as an input and make a model which would give better prediction for PM10.

In order to select the best fit model, different machine learning models were designed, trained and tested using the available data. These includes linear regression and feedforward neural networks for regression and classification (Probabilistic Neural Network), which contain the whole interval of possible values separated in high number of intervals. This is a particular case of application of classification. Aside from that, the design of the models use supervised learning, which means the output target variable is already available for the training. Different features and respectively variables were considered for the models and the performances were compared. Current (CT), CO<sub>2</sub>, Relative Humidity (RH) and Air Velocity (V) Features are the variables chosen as inputs for the model. The splitting of the measured data is set at random, accordingly, the ratio of the training and test data sets was set manually. Figure 6 shows the schematic of the Feedforward Neural Network (FFNN) Model for regression using Current (CT), CO<sub>2</sub>, Relative Humidity (RH) and Air Velocity (V) as Inputs, feeding 2 Hidden Layers with each 10 Nodes and PM10 as output.



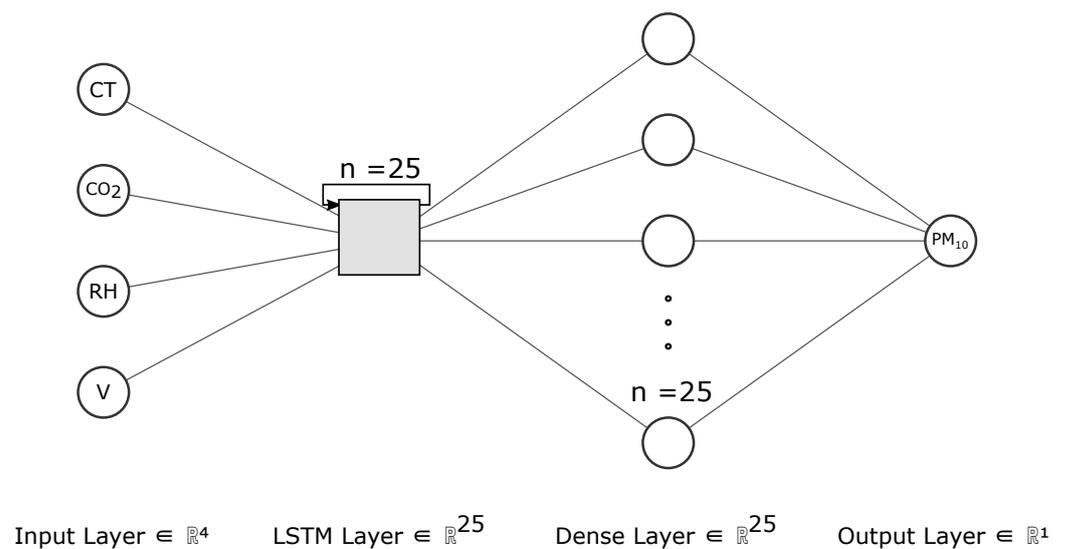
**Figure 6.** Feedforward Neural Network (FFNN) schematic with 4 inputs, 2 hidden layers @10 nodes.

The schematic of the Feedforward Neural Network Model for classification (Probabilistic Neural Network) using the same inputs, two pattern (hidden) layers and PM10 as output is shown in Figure 7. The number of nodes in the pattern layers are equal to the number of training points in the dataset. In the first pattern layer, the training data set uses 40% of the data, while the second layer utilizes 28% of the data set.



**Figure 7.** Probabilistic Neural Network (PNN) schematic with 4 inputs and 2 pattern layers.

Last, Long Short-Term Memory (LSTM), a deep learning algorithm, is used in this study, the data from the aforementioned parameters are used as input variables, 1 dense layer with 25 LSTM Blocks and the particulate matter PM10 as output variables. Figure 8 shows the schematic of the LSTM model for this study.



**Figure 8.** Long Short-Term Memory (LSTM) schematic with 4 inputs and 1 dense layer.

The models were trained with different amount of data starting from 20% to 80%. The models did not overfit until 40%, as the results continue to get better with the increase in the amount of training data. Models showed little betterment with increase in training data above 40%. The computational time increases drastically with the increase in training data above 40% for LSTM and FFNN, so 40% training data was the ideal choice to compare the models.

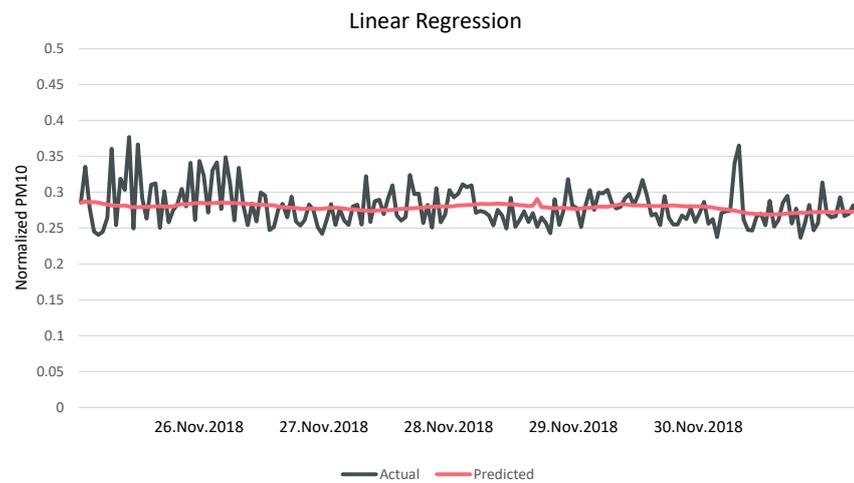
After building and training the model with the defined amount of data, the rest of the measured data was used for validating (testing) the model. The model is tested with different kinds of statistical methods such as  $R^2$  criteria for best fit, Minimum Absolute Error (MAE), Minimum Square Error (MSE) and Root Mean Square Error (RMSE). Visualization as well as statistical evaluation have been carried out to compare the actual data and the predicted data. By comparing different models with statistical as well as visual criteria, the best model for the application is chosen.

#### 3.4.1. Model Result of Linear Regression, Feedforward Neural Network and Probabilistic Neural Network

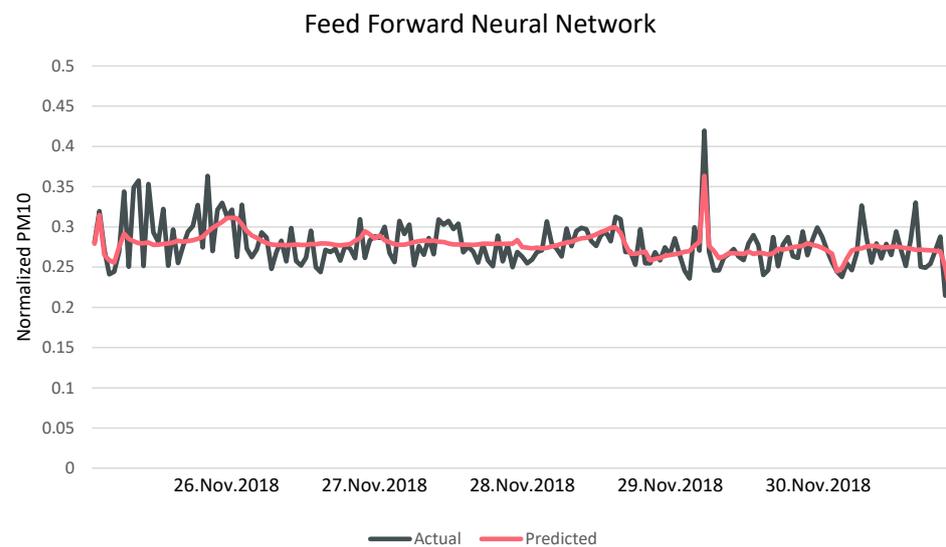
From the correlation study, it was sufficient to say that a linear relationship exists up to some extent between the input variables and desired output. Building a linear regression model using the chosen variables from the correlation study seems to be a logical way to go. After training the parameters by using the training data, which consist of 40% of the data set, the model was tested using the rest of the data set.

Figure 9 shows a direct comparison of the prediction from the linear regression model and the test data from the measured data set. An evident discrepancy exist can be seen from the comparison. Visually, Figure 10 shows a more satisfying result, which means the feedforward neural network model is a better fit of a model compared to linear regression. However, both results are not sufficient to use because of their inaccuracy. Both models use regression methods, in which the output layer consists of only one neuron and it gives a continuous output value.

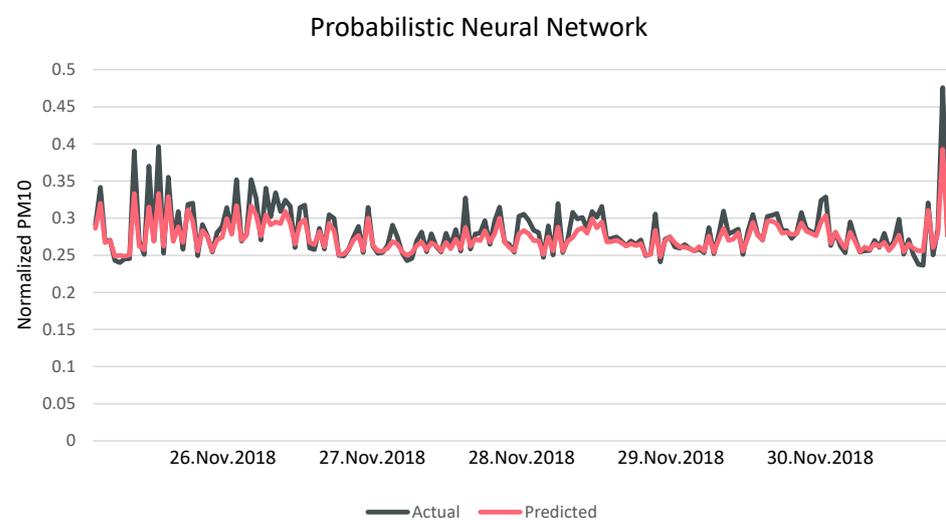
Supervised learning classification methods also serve as alternative; they are used in building the model using probabilistic Neural Network. The result, thus the comparison to the test data is shown in the Figure 11, which displays a more satisfying result in compare to Figures 9 and 10. However, merely relying on visual analysis, a scientific analysis would not be valid.



**Figure 9.** Comparison between the Linear Regression prediction and the test data.

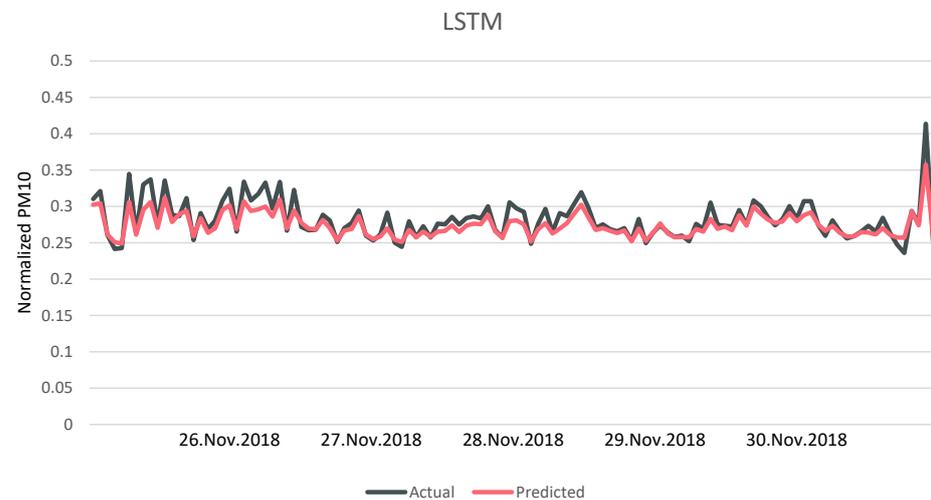


**Figure 10.** Comparison between the Feedforward Neural Network prediction (FFNN) and the test data.



**Figure 11.** Comparison between the Probabilistic Neural Network (PNN) prediction and the test data.

For example, the results shown in Figure 12, which show the visual representation of LSTM model output with the actual measured data, cannot show the difference or benefit of LSTM model compared to PNN model in Figure 11. Therefore, one needs to compare the results using statistical methods as well.



**Figure 12.** Comparison between the Long short-term memory (LSTM) prediction and the test data.

### 3.4.2. Statistical Analysis of the Models

In order to do statistical checks, the models are tested using different kinds of methods, such as  $R^2$ -criteria for best fit, Minimum Absolute Error (MAE), Minimum Square Error (MSE) and Root Mean Square Error (RMSE) and using their results the best fit model can be chosen.

Table 6 shows that the LSTM Model has the lowest error in all error checks and the best squared correlation. In contrast to the linear regression model that shows a poor fit on the  $R^2$ -Corellation and the lowest performance on the error tests, the LSTM Model shows good results, which is well matched with the visual comparison of Figures 4 and 5, and thus supports the proof that LSTM is capable to be the best fit model for this study.

**Table 6.** Results of the statistical checks of the models with Current (CT),  $CO_2$ , Relative Humidity (RH) and Air Velocity (V) as inputs.

Model	$R^2$	MAE	MSE	RMSE
Linear Regression	0.008	0.03563	0.00377	0.05689
FNN	0.211	0.02336	0.00301	0.05126
PNN	0.453	0.01901	0.00267	0.04956
LSTM	0.821	0.01020	0.00122	0.0257

In order to get a better result, one could use a cumulative training method. A cumulative learning method uses aggregation of data as it grows with time. Consequently, it uses knowledge acquired on prior training to improve learning performance on subsequent training. On the contrary, a static training method uses discrete data, which means for each new training time period the algorithms are reset and fed with new data. Fixed training data are applied to a machine learning algorithm, and it does not use any knowledge from prior training. Therefore, cumulative method reuses learned knowledge to constrain new training, whereas static method depends entirely upon new training data as external inputs [27]. Thor [28] gave further description and definition of cumulative learning in the context of machine learning in detail.

Figure 13 shows the comparison of static and cumulative training methods and the statistic result from this study according to  $R^2$  criteria. In the static training, the algorithm was trained for five days each month with new data. On the other hand, cumulative

training uses aggregated data of the current and previous months. The cumulative training method showed a gradually increasing  $R^2$  score reaching 0.81  $R^2$ , while the static method does not show a constant behavior and staying under 0.78  $R^2$ . This shows, that in training the LSTM Model, the cumulative training method is preferred.

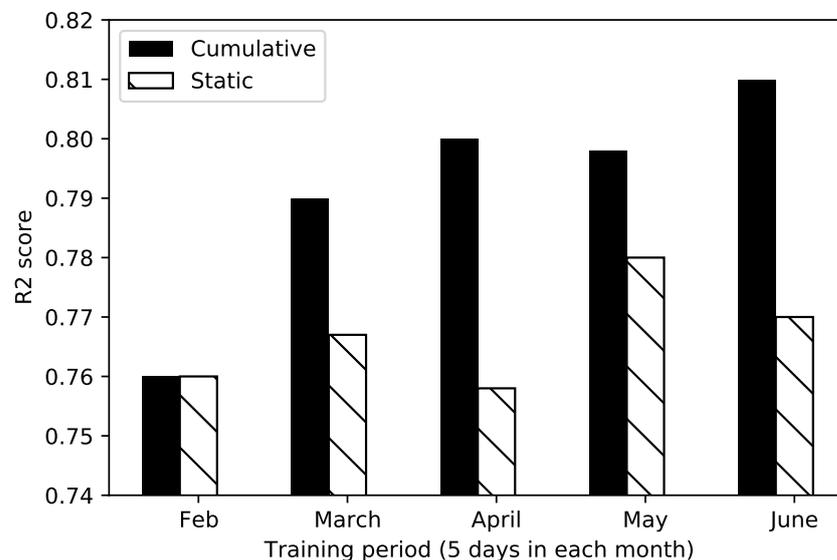


Figure 13. Long Short-Term Memory (LSTM) model analysis.

#### 4. Conclusions and Future Scope

This paper studied proof of concept for the Heating, Ventilation and Air-Conditioning (HVAC) system for a welding area. The proof of concept was modeled using four different machine learning algorithms and their performances were compared. The algorithms implemented were linear regression, feedforward neural network for regression, probabilistic neural network (Bayesian neural networks) for classification and Long Short-Term Memory (deep learning algorithm). Long Short-Term Memory showed the best result and potential for the control system.

A complete cyber-physical HVAC system in a body shop use case for the substitution of cost-intensive fine dust sensors was presented. The CPPS is then applied and validated in a real-world environment in one of the welding areas of Volkswagen's production hall in Wrzesnia, Poland. A setup was configured and established based on the regulatory threshold to collect experimental data during production hours, in order to have a solid validation ground. The data were collected in 2 phases (in summer and winter), each extended through 5 production days. The concerned data set consist of air temperature; relative humidity; air velocity; electric current used for welding; carbon dioxide; inhalable coarse particles; PM10, which are dust particles with a diameter of 10 micrometers (10  $\mu\text{m}$ ) or less and fine particles; and PM2.5, which has a diameter of 2.5  $\mu\text{m}$  or less.

These were measured in the welding area during manual welding corrections, welding of floor elements (galvanized steel) with the welding torch (MAG method), sanding and cleaning of welded seam. As a double-proofing measure, 8-hour gravimetric measurements were taken, commissioned by the Volkswagen work safety department, in order to ensure that the reductions of air exchange rate kept the exposition thresholds under the regulated limits. Gravimetric measurement (respectively analysis) describes a method to quantify a substance or chemical constituent in a mixture based on its mass. In our case, the relevant substance is fine dust and the mixture is air. The gravimetric measurements were executed according to PN EN 689:2018. It showed that the reduction of air exchange rate does not have a linear correlation with the change in the amount of fine dust in the working area. The correlation study according to Spearman's rho and Pearson showed that only four variables have direct correlation with the outputs (PM10 and PM2.5): relative humidity, CO<sub>2</sub> Concentration, electric current and air velocity in the area. Because PM10 and PM2.5

show a strong increasing correlation with each other, PM<sub>2.5</sub> could be excluded as an output variable. This will save a generous amount of computing time and cost. As supervised machine learning and deep learning algorithms are used in this study, the data from the aforementioned parameters are used as input variables and the particulate matter as output variables. The data were then split for training and testing of the models. The models are then tested statistically using different methods, such as  $R^2$  criteria for best fit, Minimum Absolute Error (MAE), Minimum Square Error (MSE) and Root Mean Square Error (RMSE). This shows that LSTM dominated the test with 0.821 discrepancy on  $R^2$ , 0.01020 on MAE, 0.00122 on MSE and 0.0257 on RMSE. As a comparison, the Linear Regression Model showed a poor result with 0.008 discrepancy on  $R^2$ , 0.03563 on MAE, 0.00377 on MSE and 0.05689 on RMSE. The results were also examined using graphical method. The LSTM or Long Short-Term Memory showed the best result and therefore is best suited for the HVAC control concept.

In the future it is possible to enhance the HVAC control using the LSTM model. It could be possible to install several micro controllers in the factory hall, that send the acquired data that is not expensive to measure data, such as relative humidity, CO<sub>2</sub>, electricity current and air velocity. Data can then be used to train the LSTM model continuously. The output of the models can send computed fine dust values in real time to the HVAC system, which would be used as control parameter, besides typical control parameters, such as desired temperature, relative humidity and CO<sub>2</sub> especially on factory scale, the cost savings would be significant and are very interesting for factory operators and planners.

**Author Contributions:** Y.W.: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing—Original draft, Visualization, Project administration. M.V.: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing—Reviewing and Editing. R.K.: Methodology, Software, Visualization, Investigation, Data Curation. C.F.: Supervision. C.H.: Supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research regarding the presented use case in this paper was funded by the Federal Ministry for Economic Affairs (BMWi) by means of the 7th Energy Research Programme of the German Federal Government under grant number 03ET1660A (3DEMO—Safe and energy efficient factories through 3D emission monitoring).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data was obtained from Volkswagen AG and are available from the authors with the permission of Volkswagen AG.

**Conflicts of Interest:** The authors declare no conflicts of interest. The results, opinions and conclusions expressed in this paper are not necessarily those of Volkswagen AG.

## References

1. Hyoungh, S.K.; Ju, Y.L.; Sang, S.C.; Hyun, K.; Jun, H.P.; Ji, Y.S.; Bo, H.K.; Sang, D.N. Smart manufacturing: Past research, present findings, and future directions. *Int. J. Precis. Eng. Manuf.-Green Technol.* **2016**, *3*, 111–128.
2. Thiede, S. Smart Environmental Sustainability of Cyber Physical Production Systems. *Procedia CIRP* **2018**, *69*, 644–649; ISSN 22128271. [CrossRef]
3. IEA's World Energy Balances and Statistics. Total Final Consumption by Source (TFC), World, Data Retrieved from IEA World Energy Balances and Statistics. 2020. Available online: <https://www.iea.org/data-and-statistics?country=WORLD&fuel=Energy%20consumption&indicator=TFCbySource> (accessed on 29 October 2021).
4. IEA's World Energy Balances and Statistics. Share of Total Final Consumption (TFC) by Sector, World, Data Retrieved from IEA World Energy Balances and Statistics. 2020. Available online: <https://www.iea.org/data-and-statistics?country=WORLD&fuel=Energy%20consumption&indicator=TFCbySector> (accessed on 29 October 2021).
5. IEA's World Energy Balances and Statistics. Total Final Consumption (TFC) by Sector, World, Data Retrieved from IEA World Energy Balances and Statistics. 2020. Available online: <https://www.iea.org/data-and-statistics?country=WORLD&fuel=Energy%20consumption&indicator=TFCShareBySector> (accessed on 29 October 2021).
6. Volkswagen, A.G. *Energieaufteilung Nach Gewerke Volkswagen*; 2018, Unpublished internal company document.

7. Wang, S.K. *Handbook of Air Conditioning and Refrigeration*; Engineering Handbook; McGraw-Hill Education: New York, NY, USA; Vulkan-Verlag GmbH: Essen, Germany; DIV Deutscher Industrieverlag GmbH: München, Germany, 2015; ISBN 97800706816752001. Available online: <https://books.google.de/books?id=0tVSAAAAMAAJ> (accessed on 29 October 2021).
8. Recknagel, H. *Heizung + Klimatechnik mit Neuem Hauptkapitel "Energiekonzepte"*; Vulkan-Verlag GmbH, DIV Deutscher Industrieverlag GmbH: Essen, München, 2011. Available online: [http://slubdd.de/katalog?TN\\_libero\\_mab216556880](http://slubdd.de/katalog?TN_libero_mab216556880) (accessed on 29 October 2021).
9. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA). TRGS 900, Technische Regeln für Gefahrstoffe—Arbeitsplatzgrenzwerte in Bundesarbeitsblatt (BArBl.) Heft 1/2006, S. 41–55. Bundesministerium für Arbeit und Soziales (BMAS). Januar 2006, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Friedrich-Henkel-Weg 1–25. D-44149 Dortmund. Available online: [https://www.baua.de/DE/Angebote/Rechtstexte-und-Technische-Regeln/Regelwerk/TRGS/pdf/TRGS-900.pdf?\\_\\_blob=publicationFile](https://www.baua.de/DE/Angebote/Rechtstexte-und-Technische-Regeln/Regelwerk/TRGS/pdf/TRGS-900.pdf?__blob=publicationFile) (accessed on 29 October 2021).
10. Födisch Umweltmesstechnik, A.G. Feinstaubsensor FDS 15 Produktinformation, Zwenkauer Straße 159, 04420 Markranstädt, Deutschland. 2020. Available online: [https://www.foedisch.de/app/download/12654990627/de.FDS%2015\\_Produktinformation.pdf?t=1596630054](https://www.foedisch.de/app/download/12654990627/de.FDS%2015_Produktinformation.pdf?t=1596630054) (accessed on 29 October 2021).
11. Födisch Umweltmesstechnik, A.G. *Feinstaubsensor Offizielles Angebot*; Zwenkauer Straße 159, 04420 Markranstädt, Deutschland, 2018. Unpublished internal company document.
12. Al-Fuqaha, A.; Guizani, M.; Mohammadi, M.; Aledhari, M.; Ayyash, M. Internet of Things: A Survey on Enabling Technologies, Protocols and Applications. *IEEE Commun. Surv. Tutor. J.* **2015**, *17*, 2347–2376. [[CrossRef](#)]
13. Bajrami, X.; Murturi, I. An efficient approach to monitoring environmental conditions using a wireless sensor network and NodeMCU. *Elektrotech. Inftech.* **2018**, *135*, 294–301. [[CrossRef](#)]
14. Li, L.; Zhang, R.; Sun, J.; He, Q.; Kong, L.; Liu, X. Monitoring and prediction of dust concentration in an open-pit mine using a deep-learning algorithm. *J. Environ. Health Sci. Eng.* **2021**, *19*, 401–414. doi:10.1007/s40201-021-00613-0. [[CrossRef](#)] [[PubMed](#)]
15. Umweltbundesamt, Hintergrundpapier zum Thema Feinstaubbelastung in Deutschland, Wörlitzer Platz 1, 06844 Dessau-Roßlau. 2009. Available online: <https://www.umweltbundesamt.de/sites/default/files/medien/publikation/long/3565.pdf> (accessed on 29 October 2021).
16. European Environment Agency, Sources of Air Pollution in Europe, Published 13 Jun 2014; Last modified 23 Nov 2020. Available online: [https://www.eea.europa.eu/ds\\_resolveuid/INF-13-en](https://www.eea.europa.eu/ds_resolveuid/INF-13-en) (accessed on 29 October 2021).
17. Madanchi, N.; Leiden, A.; Winter, M.; Asbach, C.; Lindermann, J.; Herrmann, C.; Thiede, S. Cutting fluid emissions in grinding processes: Influence of process parameters on particle size and mass concentration. *Int. J. Adv. Manuf. Technol.* **2019**, *101*, 773–783. [[CrossRef](#)]
18. European Environment Agency, 6.4. Suspended particulates (TSP/SPM), Last Modified 23 Nov 2020. Available online: <https://www.eea.europa.eu/publications/2-9167-057-X/page021.html> (accessed on 29 October 2021).
19. Xing, Y.F.; Xu, Y.H.; Shi, M.H.; Lian, Y.X. The impact of PM2.5 on the human respiratory system. *J. Thorac. Dis.* **2016**, *8*, E69–E74. [[CrossRef](#)] [[PubMed](#)]
20. Scott, R.M. *Chemical Hazards in the Workplace*; CRC Press: Boca Raton, FL, USA, 2020; ISBN 9781000716344. Available online: <https://books.google.de/books?id=CkQPEAAAQBAJ> (accessed on 29 October 2021).
21. Riccelli, M.G.; Goldoni, M.; Poli, D.; Mozzoni, P.; Cavallo, D.; Corradi, M. Welding Fumes, a Risk Factor for Lung Diseases. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2552. [[CrossRef](#)] [[PubMed](#)]
22. Spiegel-Ciobanu V.E. Aktuelles aus Gesetzgebung und Regelwerk zu Schadstoffen beim Schweißen. In *Schneiden und Schweißen, Ausgabe 04/2016*; BGHM—Berufsgenossenschaft Holz und Metall: Mainz, Germany, 2016. Available online: [https://www.esta.com/fileadmin/mount/ESTA\\_DE/Download/Wissenswertes/BGHM-Regelwerk-Schadstoffe-Schweissen.pdf](https://www.esta.com/fileadmin/mount/ESTA_DE/Download/Wissenswertes/BGHM-Regelwerk-Schadstoffe-Schweissen.pdf) (accessed on 29 October 2021).
23. Deutsche Institut für Normung e.V. (Hrsg.). *DIN EN 689, Exposition am Arbeitsplatz—Messung der Exposition Durch Einatmung Chemischer Arbeitsstoffe—Strategie zur Überprüfung der Einhaltung von Arbeitsplatzgrenzwerten; Workplace Exposure—Measurement of Exposure by Inhalation to Chemical Agents—Strategy for Testing Compliance with Occupational Exposure Limit Values*; Beuth Verlag GmbH: Berlin, Germany, 2020.
24. Polish Committee for Standardization. *PN-EN 689, Exposure at Work Stations—Measurements of Inhalation Exposure to Chemical Agents—Strategy for Testing Compliance with Limit Values*; Polish Committee for Standardization: Warsaw, Poland; 2019.
25. Bishop, C.M. *Pattern Recognition and Machine Learning, Information Science and Statistics*; Softcover Published in 2016; Springer: New York, NY, USA, 2006. Available online: <https://cds.cern.ch/record/998831> (accessed on 29 October 2021).
26. Spearman Rank Correlation Coefficient. In *The Concise Encyclopedia of Statistics*; Springer: New York, NY, USA, 2008; doi:10.1007/978-0-387-32833-1\_379. [[CrossRef](#)]
27. Michelucci, P.; Oblinger, D.; Sammut, C.; Webb, G.I. (Eds.) *Cumulative Learning, Encyclopedia of Machine Learning and Data Mining*; Springer: Boston, MA, USA, 2017; pp. 306–314. ISBN 978-1-4899-7687-1. [[CrossRef](#)]
28. Thórisson, K.; Bieger, J.; Li, X.; Wang, P. Cumulative Learning. In *International Conference on Artificial General Intelligence*; Springer: Cham, Switzerland, 2019; Volume 7, pp. 198–208. ISBN 978-3-030-27004-9. [[CrossRef](#)]