1st Workshop on Low-Cost Digital Solutions for Industrial Automation

Virtual sensor architecture for indoor air quality monitoring

Jorge Merino[1*], Daiki Ikeuchi[2], Nicola Moretti[1], Anandarup Mukerjee[1], Luning Li[1], Stylianos Karatzas[1], Sebastian Pattinson[2], Ajith Parlikad[1]

[1] Distributed Information and Automation Laboratory, Department of Engineering, University of Cambridge, Institute for Manufacturing, Alan Reece Building, 17 Babbage Road, Cambridge, United Kingdom
[2] Computer Aided Manufacturing Laboratory, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom
[*]jm2210@cam.ac.uk

Abstract: Human exposure to poor air quality is a leading risk factor in the Global Burden of Disease (GBD) study, estimating 22,000 premature deaths related to indoor air pollution in 2019 in Europe. Diverse pollutants are found in manufacturing environments resulting from both combustion and non-combustion sources, including Particulate Matter and Volatile Organic Compound. Internet of Things (IoT) air quality monitoring can enhance awareness and support informed decision making towards better air quality. However, hardware sensors are not always capable of monitoring particular characteristics and behaviour of a pollutant, for instance, spatial limitations may impede deploying sensors close enough to the source of the pollutant. Virtual Sensors can extend hardware sensing options via signal processing and data integration. This paper presents an architecture for training and deploying virtual sensors. A virtual sensor is implemented using the architecture in the context of additive manufacturing to estimate the production of Volatile Organic Compounds (VOCs) of 3D printers and their transfer into the rest of the space. In the case study, the 3D printers are installed inside cabinets to limit the transfer of pollutants to the exterior. Several of these virtual sensors are deployed to monitor the VOCs produced by the 3D printers and the transfer rate out of the cabinets. The paper includes some early results and initial insights on the accuracy and usefulness of virtual sensors. Virtual Sensors can be cost-effective solutions when monitoring systems are escalated by reducing number of hardware sensors and complexity.

1. Introduction

Poor air quality in manufacturing environments has been associated with work-related illnesses like lung diseases [1]. The World Health Organization (WHO) reported that poor air quality caused 4.2 million deaths in 2016, of which, primarily, 17% were due to strokes, 25% were due to chronic obstructive pulmonary diseases (COPD), and 26% were due to respiratory diseases [2]. The Global Burden of Disease (GBD) study estimated that ambient air pollution causes almost 4.1 million premature deaths, and over 200 million disability-adjusted life-years (DALYs) globally each year [3], 22,000 of those premature deaths in 2019 in Europe were related to combustion of solid fuels. Diverse pollutants are found in manufacturing environments resulting from both combustion and non-combustion sources. The most dangerous indoor air pollutants include Particulate Matter (e.g., PM 2.5, 5 and 10), Oxides (e.g., NOx, O3, CO, SO2), volatile and semi-Volatile Organic Compounds (VOCs), radon, toxic metals, and microorganisms [4], [5].

Pollutants generated during manufacturing processes depend on the methods and materials used [6]. For instance, in additive manufacturing (i.e., building three-dimensional solid objects from their digital models by selectively accumulating material layer-by-layer [7]), the main methods are fused deposition modelling (FDM), powder bed fusion (PBS), inkjet printing and contour crafting, stereolithography (SLA), and laminated object manufacturing (LOM). The most popular materials are metals and alloys, polymers and composites, ceramics, and concrete [8]–[13].

Air quality monitoring enhances awareness and support informed decision making towards better air quality. Research mostly focuses on high-end monitoring [5], [6], [14]–[19]. Some authors try to make low-cost monitoring feasible by integrating it with signal processing and analysis [17], [20].

Pre-processing and data analytics are used to tackle known low-cost sensor problems like predicting missing and dependent variables, estimating missing data, cleaning noisy data and outliers, and pattern recognition. Some authors have automated these techniques via software sensors (or soft-sensors) to extend the capabilities of hardware sensors. Soft-sensors are online instruments of measurement that supplement monitoring [21], [22]. Soft-sensors are combination models using known process data to predict missing measurements and reconstruct missing variables from available measurements [23]. Soft-sensors have been used recently for drift correction of NOx emissions in industrial settings [24], and quality of products and tools in various manufacturing environments with chemical or mechanical processes [25]–[28]. Some authors also differentiate between soft and virtual sensors, depending on whether the estimations are supported by online measurements by other monitoring entities (i.e., virtual sensor) or not (i.e., soft-sensor) [29]. [30] defines a virtual sensor as "a type of software that processes what a physical sensor otherwise would. It learns to interpret the relationships between the different variables and observes

https://doi.org/10.1049/icp.2023.1749

readings from different instruments.” [31] reviews the literature around virtual sensing in buildings and summarises main applications of virtual sensors, including zone-level air filtration rate and heat, room level occupancy, return heating water temperature and flow rate estimation, supply air temperature and flow rate estimation, and heating and power consumption.

In terms of indoor air quality, soft and virtual sensors have been used to extend the capabilities of hardware sensors as well as for reducing the number of sensors required and monitor targets that out of reach (i.e., it is not possible to install a hardware sensor). [32] trained Neural Network virtual sensors on data from fixed air quality stations located in urban areas to estimate NO, NO2, NOx, SO2, O3, CO, Cd, PM2.5, and PM10 concentration and virtually recreate the pollution in entire neighbourhoods. [33] implemented a virtual sensor for selective detection of hazardous VOCs using a gas sensor array. [34] developed a PM estimator for urban areas based on static and mobile sensors. [35] developed soft sensors to predict and monitor PM2.5 in an underground metro system using just-in-time (JIT) learning. [36] evaluates the prediction performance of a composite model integrating latent variables of kernel partial least squares with relevance vector machine (KPLS-RVM) against least squares support vector machines (LSSVM) and subvariants to improve the prediction performance of conventional soft sensors, and test this approach in the aforementioned subway system and to estimate chemical oxygen demand (COD) in a wastewater treatment process (WWTP). [37], [38] designed a soft-sensor to predict indoor PM2.5 using neural circuit policies (NCP), and continuous recurrent neural networks (RNN). [39] proposes a soft sensor design based on image capturing and deep learning analysis of bio indicators from natural monitoring specimens like cryptograms, and bryophytes which are sensitive to pollutants. [40] suggest a methodology for air quality virtual sensor development and calibration using reference equipment.

While most of these solutions for virtual and soft-sensors have been applied to estimate pollutants both in indoor and outdoor environments, there is a gap on research on air quality in manufacturing environments using virtual or soft-sensors. High-end equipment has been vastly used to monitor manufacturing environments, and soft-sensors have been developed to estimate product and tools quality in manufacturing, but not air quality. This paper proposes an architecture streamline training and deployment virtual and soft sensors in a scalable manner. Additionally, three virtual sensors based on linear methods to estimate VOC in an additive manufacturing environment with three 3D printers enclosed in cabinets with extraction to estimate the transfer rate from inside the cabinet to the rest of the environment.

The rest of the paper is structured as follows. Section 2 introduces the key related works in air quality monitoring in manufacturing, and virtual sensing. Section 3 describes the virtual sensing method for air quality monitoring, including the architecture and the models used. Section 4 displays a case study where the virtual sensing method was implemented, including the setting, the experiments setup and the results obtained. Section 5 reaches depicts some conclusions out of this research.

2. Related work

This section summarises some related works in the literature, including alternatives for air quality monitoring in manufacturing, and techniques for soft and virtual sensors.

2.1. Air quality monitoring in manufacturing

There are two trends related to air quality monitoring in manufacturing: high-end and low-cost monitoring. The former uses addressed chemical techniques enabled by accurate equipment, which in turn comes at a cost and often enable one-off sampling. The latter uses off-the-shelf generic sensors produced for all kind of pollutants in a group (e.g., all types of VOCs). The techniques used for low-cost sensors do not focus on a particular chemical reaction, but rather on alternatives that ensure a longer life of the sensor for continuous monitoring.

In the first group, [19] used a range of equipment, including sampling pumps, filters, electrostatic classifiers, particle counters, and gas detectors to monitor PM and VOC emissions from an inkjet 3D printer. [6] uses a membrane pump directly connected to a chamber through an extraction tube and an activated carbon vial on PLA and ABS to monitor VOCs. [14] uses absorption tubes, an air sampling pump, a thermo-desorption instrument, a chromatography system, and a mass spectrometer to analyse the VOC produced by diverse techniques for 3D printing in six different settings. [15] uses a scanning mobility particle sizer with a differential mobility analyser to measure particle size distribution to identify ultrafine particles (UFP) in a FDM 3D printing process with different materials. [16], [18] conducts a similar experiment monitoring ultrafine particles produced from metals in a Powder Bed Fusion 3D printing process.

In the second group, [20] monitors PM2.5 with SPS30 sensors in a multi-sensor node placed in the ventilation system of an incremental manufacturing lab combining additive and subtractive manufacturing.

2.2. Methods used in Soft and Virtual sensor development

Techniques used are multifaceted and can be classified into statistical, analytical, probabilistic, and machine learning. Statistical methods include techniques like regression, principal component analysis (PCA) and partial least squares (also known as projection to latent structures) (PLS), Kalman filters, and modal analysis; Analytical methods comprise data scaling, correlation, scheffe tests, and drift identification; probabilistic methods encompass distribution analysis, fuzzy theory and data imputation; and Machine Learning methods involves support vector machines (SVM) and regression (SVR), long short-term memory (LSTM, auto-encoders, perceptron, which evolved into Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN), decision trees, and reinforcement learning (RL) among many others. Table 1 summarises these methods in the literature.

Additionally, some authors have suggested virtual sensors architectures [29], [31], [33], [40], [41].

Table 1. Soft-sensors, Data pre-processing and analysis methods in the literature

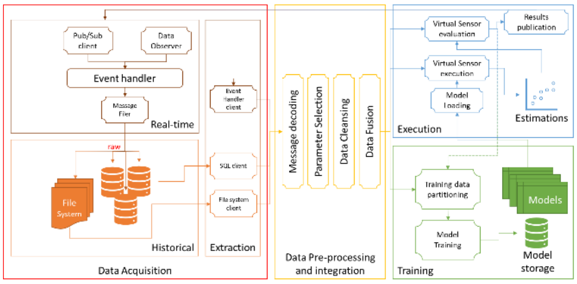
Method category	Description and common types
Statistical	Regression (Linear, Logistic, Polynomial, etc.)[42], Principal Component Analysis (PCA) and Regression (PCR) [43], [44], Partial Least Squares / Projection to latent structures (PLS) [45], [46], Linear Quadratic Estimation (LQE) / Kalman filter [47], [48], Modal analysis [49], [50], Stepwise regression [51], Iterative algorithms [52]
Analytical	Data scaling [53], Correlation, Scheffe test [54], [55], Drift identification, Hidden context [56]
Probabilistic	Distributions [57], [58], Fuzzy methods [59]–[61], Data Imputation [41], [61]
Machine Learning	Support Vector Machine (SVM) [62], Long Short-Term Memory (LSTM), Auto-Encoders, perceptron, Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs) [63], [64], Generative Adversarial Networks [65], Decision Trees [66]–[68], Genetic Programming [69], [70], Self-organising maps [71], [72], Reinforcement Learning (RL) [64]

3. Virtual sensors for indoor air quality monitoring

3.1. *Architecture*

The architecture proposed in this paper is designed to enable both training and deployment of virtual and soft sensors. Fig. 1. *Architecture for Virtual and Soft-sensor training and deployment*Fig. 1 shows the architecture, which is modularised into data acquisition, data pre-processing and integration, model training, and execution.

Fig. 1. *Architecture for Virtual and Soft-sensor training and deployment*



Data acquisition is partitioned into real-time and historical pipelines. The real-time pipeline enables two types of agents to acquire data: Pub/Sub clients that subscribe publishers of data (e.g., MQTT, Kafka, WebSocket), and data observers actively query new data from non-publish subscribe model sources in a predetermined frequency basis (e.g., SQL data observer for a time-series database every 15 minutes). Both types of agents publish new data in the event handler which follows a publish-subscribe model. The agents also attach metadata (e.g., timestamps on data arrival, data source) to the original payload. Message filers are responsible for storing data into a file system or databases. Data is stored on its raw format from all sources including attached metadata. New data becomes part of the historical data pool. A data extraction module is responsible for accessing either the real-time or the historical pipelines for model execution and training, respectively. Data extraction in real-time subscribes to the event handler and passes on the new incoming readings; data extraction of historical data retrieves readings stored according to a query that includes, a time frame and required targets (i.e., monitoring devices).

Data pre-processing and integration is pipelined into four steps. The first module decodes individual readings when necessary. Often sensor signals are encoded in bytes rather in human-readable messages, therefore, this first module comprises decoders for all monitoring sources. The module is capable of identifying the source of the reading thanks to the attached metadata and applying the corresponding decoder. Second module helps with parameter selection based on the virtual sensor requirements. It filters the required parameter in every reading. Third module is designed to cleanse data individually in each selected parameter. Some methods for data cleansing are outliers' identification and deletion, noise reduction, modal analysis, or drift identification. Fourth module conducts data fusion. When working with multiple monitoring devices, timestamps of the readings are not synchronised, even if the devices have been synchronised. This creates gaps in the data when it needs to be integrated. In case several sources are required by a virtual sensor to make estimations, this module needs to wait until all readings are ready (or discard them using a timeout). Some methods that help with this problem are case deletion, and data imputation. Techniques for feature selection and extraction like PCA, PLS, transfer functions, wavelets, Hampfel filters can also be applied at this point. Prepared and integrated data is available for training new models as well as supporting the estimation of pre-trained models.

Training and Execution module are inspired by the virtual sensing design proposed in [41], but have been redesigned to accommodate automated calibration. The training module queries required data (i.e., desired parameters and features within a timeframe) through the previews modules. Data is partitioned into training and testing datasets, and it is fed into the models selected for training (e.g., ANN, SVM, linear models). Since data partitioning and training is executed in separate modules, the same datasets can be used to train different models. Each model is then stored and ready for deployment.

The execution module starts by loading an available pre-trained model. It implements an observer function to load new versions of the model when available in the model storage. The virtual sensor execution module subscribes to required data that matches the data sources, parameters, and features that the loaded model was trained on. For every reading received, an estimation is produced. In case the virtual sensor is executed as a soft-sensor, it requires a sampling-rate configuration for making estimations without the support of real-time readings from any other source. In either case, new estimations produced, which are delivered to a sensor evaluation module. The evaluation of a virtual sensor is based on different indicators, for instance, the accuracy of the model recorded at the time of training, the comparison between the distribution of training data and operational data, the time to produce an estimation, or time since the last calibration of the virtual sensor. The combination of these indicators results into the uncertainty of the estimations. Additionally, the virtual sensor evaluation module can trigger a request to the training module to re-train (i.e., re-calibrate) a particular model when uncertainty is high. Both the estimation and the uncertainty are published through the result publication module, which attaches additional metadata to every estimation (e.g., virtual sensor id, time of estimation, models used, data source timestamps, etc.). Thus, the virtual sensor behaves seamlessly as any other monitoring device, and estimations produced can be used by other virtual sensors and applications. It is important to highlight that the execution module in this architecture can be deployed in different system as long as it is connected to the other modules (e.g., in the edge).

4. Case study: 3D printing lab

This section reports a case study where the architecture was implemented to develop three virtual sensors using linear models to estimate VOCs. The case study includes the description of the context and setting of the environment, the design process of the virtual sensors implementing the architecture, experiments conducted to collect data, and the validation of the virtual sensors.

4.1. Context and setting

The context of the case study is an additive manufacturing laboratory that uses 3D printers to produce parts. 3D printers are either enclosed in cabinet made of methacrylate panels with windows for accessibility or in stations in the rest of the environment.

Fig. 2 shows the setting for this case study which includes three 3D printers enclosed in one of these cabinets with two double door windows and two sliding windows (one of which never opens), and an extraction tube at the top.

Fig. 2. *Case study setting. Cabinet with three 3D printers*

3D printers use the Fused Deposition Modelling (FDM) technique in which a heated nozzle melts a solid thermoplastic filament and deposits multiple thin layers of extruded plastic to form a solid three-dimensional shape. Materials used in the experiments are Acrylonitrile Butadiene Styrene (ABS), Polylactic Acid (PLA), and thermoplastic polyurethane (TPU). According to other studies [5], [6], [9], [15], during the melting process, these materials emit Volatile Organic Compounds (VOCs) and Particulate Matter (PM). These studies highlight that ABS is more toxic than PLA, but that PLA is not free from harmful emissions to health issues when melted at temperatures over 200 °C.

4.2. Virtual Sensors

This subsection describes the design of the virtual sensors following the architecture.

For data acquisition, a single pub/sub client subscribes to data from the MQTT broker. Depending on the topic data is published into, it attaches metadata to the payload on arrival timestamp and source of data. Data is passed through the event handler and stored into the file system in JSON format. Every message (including the raw payload and metadata) is compiled into a day-level JSON file. During extraction, an MQTT client subscribe to required data from the virtual sensors during execution and relays messages from the event handler in real-time; a file system client queries the JSON files to extract the timeframe and variables required by the virtual sensors for training.

During data pre-processing and integration, every message is decoded using Synetica, DFRobot, and Monnit decoders. Required parameters from the virtual sensors are filtered in the parameter selection module. For data cleansing, the main methods used are for outliers' identification and deletion based on zscores. Data fusion module implements synchronisation based on a timeout set based on the sampling rate of the sensors selected. Timeout is set to 30 seconds since it the maximum sampling rate of the hardware sensors (i.e., Synetica sensors). Timeout starts when a reading from one of the required sensors arrives, and if all readings from all required sensors are not ready before the timeout, they are discarded based on arrival time.

Additionally, two methods are applied for data fusion: case deletion and data imputation. Case deletion is applied for training requests since virtual sensors must be trained on clean non-estimated data. Data imputation based on single device parameter distribution is applied for execution. This means that the distribution of every parameter from each data source is analyse individually and gaps are estimated based on it.

During training, data is partitioned in an 80%/20% proportion for all models, and same data is used to train all models initially, which enables benchmarking. Three Sci-kit learn linear models are used for VOC estimation, particularly Lasso regression, Bayesian Ridge, and Elastic Networks [73]–[75]. The goal of these models is to estimate the concentration of VOCs outside of the cabinet, based on the concentration inside of it, therefore, VOCs readings inside of the cabinet are combined with the status of the windows and the operating condition (humidity, temperature, pressure) inside of the cabinet is used to estimate VOCs concentration outside of the cabinet. Readings from the sensor on top is used for training and validation of the virtual sensors, but not for online estimation during execution. Trained models are stored in the filesystem using skops [76].

During execution, persisted models are loaded, and virtual sensors subscribe to real-time readings. The readings from inside of the cabinet were used to provide online estimations of VOC concentration outside of the cabinet. At this point of the research, the evaluation is only made based on model accuracy during training, and time since the last re-calibration of the virtual sensor, providing it as metadata of every estimation. Re-training is not yet triggered. Every estimation is published including metadata on the virtual sensor (e.g., sensor id, model used, time since last calibration).

Data requirements for the virtual sensors both during training and execution are provided in configuration files which are read by a virtual sensor factory module to automate the process.

4.3. Experiments

Some experiments were conducted in a controlled manner to generate data for sensor training. The experiment variables are described in Table 2.

Table 2. Experiment variables

Experiment Variable	**Description**
Windows	Whether the windows are open or closed. Monitored by open closed sensors.
Extraction	Extraction of the cabinet. Can be on/off. Manually adjusted.
# printers	Number of printers operating
Material	Type of material used throughout the printing. Can be more or less toxic. Mainly 3 types of materials used: TPU, PLA+ and ABS. Same manufacturer and model of each material is used in all the experiments.
Volumetric flowrate	The amount of filament used. It is dependent of the temperature of the nozzle, the melting point of the material and the filament procurement speed (filament drums can be pushed and pulled into the nozzle).
Nozzle temperature	The temperature of the nozzle of the printer. The nozzle is the component that melts the filament and moves around the bed to print geometries
Fan speed	Speed of the fan that cools the geometry down
Bed temperature	The temperature of the bed where the geometries stand while being printed

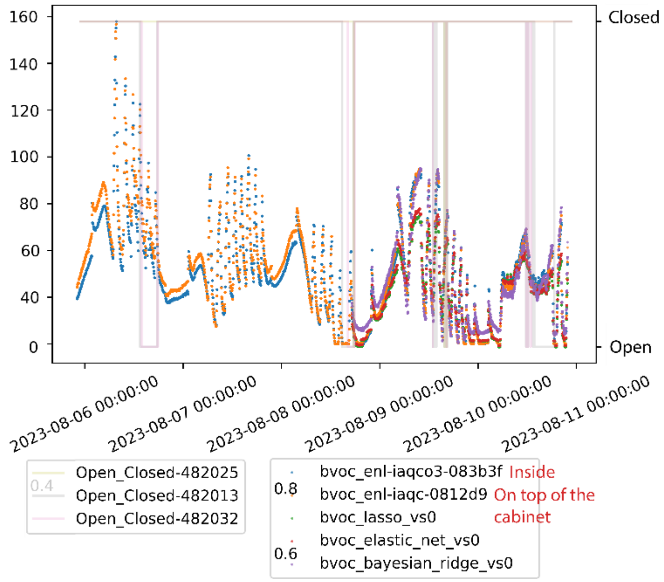
Additionally, some experiment control variables were used to characterize the experiments, including the start and end time of the printing jobs, the mass, volume, and density of the material used, and the geometry printed.

Experiments include baseline identification (i.e., no printing every window open) and cases monitoring printing jobs altering the experiment variables.

4.4. Evaluation

During evaluation, the same sensor used for the target estimation during training is used to compare with the estimations provided by the sensors. Fig. 3 shows the juxtaposition of the readings from the open closed sensors in the windows, one of the sensors inside of the cabinet, the sensor on top, and the estimations by the three virtual sensors using Lasso, Elastic networks, and Bayesian ridge regression models. The superposition of the readings demonstrates a good accuracy of the estimations. The Pearson correlation of the estimations with the readings from the sensor on top of the cabinet are for 0.965663 Lasso, 0.968593 for Elastic Networks, and 0.951995 for Bayesian Ridge. Further evaluation is deemed necessary using more data over time.

Fig. 3. Virtual Sensors estimations against real-time data



5. Conclusion

Indoor air quality monitoring in additive manufacturing environments is still in its infancy. New research is mainly focused on the use of high-end monitoring solutions to evaluate air quality in such environments. Virtual and soft-sensing approaches have demonstrated good estimates of product quality parameters in manufacturing, and urban air quality in other domains.

This paper has tackled the problem of estimating air quality using virtual sensors in an additive manufacturing environment. A modular architecture for scalable development of virtual and soft-sensors is suggested. The architecture is implemented to develop virtual sensors for monitoring VOCs in an additive manufacturing environment. This research acknowledges that high-end equipment may not always be available or even affordable. Therefore, same sensors used for supporting real-time estimations are used for validation. Despite the limitations during the real-time virtual sensor evaluation, estimations provided from three virtual sensors trained using linear models demonstrated to be fairly accurate when compared to the validation sensor. Further evaluation is necessary. Virtual sensors developed using the architecture should be evaluated using more data, and the module for online evaluation of estimations during execution of virtual sensors must be explored further to enable just-in-time re-calibration.

In addition to scalability, this approach can be used to reduce the number of sensors necessary to monitor any environment. The cost of this solution is mainly on the implementation of the architecture. Virtual sensors can be deployed in a centralised system or in the edge.

6. Acknowledgments

This work was supported by the UK's Department for Digital, Culture, Media and Sport through the project entitled “Digital Twins Enabled Indoor Air Quality Management for Healthy Living-TwinAIR”. TwinAIR is funded by UK Research and Innovation (UKRI) through the Horizon Europe Guarantee Scheme for European Union’s Horizon Europe Research and Innovation Programme, grant number 101057779, HORIZON-HLTH-2021-ENVHLTH-02-02.

7. References

[1] K. Jenab, S. M. Seyedhosseini, S. Khoury, and A. Sarfaraz, ‘An intelligent air quality monitoring model in manufacturing’, *Clean Techn Environ Policy*, vol. 14, no. 5, pp. 917–923, Oct. 2012, doi: 10.1007/s10098-012-0467-4.

[2] World Health Organization, ‘Ambient air pollution: a global assessment of exposure and burden of disease’. World Health Organization, 2016.

[3] ‘GBD Results’, *Institute for Health Metrics and Evaluation*. https://vizhub.healthdata.org/gbd-results (accessed Aug. 09, 2023).

[4] V. V. Tran, D. Park, and Y.-C. Lee, ‘Indoor Air Pollution, Related Human Diseases, and Recent Trends in the Control and Improvement of Indoor Air Quality’, *IJERPH*, vol. 17, no. 8, p. 2927, Apr. 2020, doi: 10.3390/ijerph17082927.

[5] S. Khaki, M. Rio, and P. Marin, ‘Monitoring Indoor Air Quality in Additive Manufacturing environment’, *Procedia CIRP*, vol. 90, pp. 455–460, 2020, doi: 10.1016/j.procir.2020.01.113.

[6] L. Bravi, F. Murmura, and G. Santos, ‘Additive Manufacturing: Possible Problems with Indoor Air Quality’, *Procedia Manufacturing*, vol. 41, pp. 952–959, 2019, doi: 10.1016/j.promfg.2019.10.020.

[7] ‘Additive Manufacturing Glossary’. https://www.sme.org/additive-manufacturing-glossary (accessed Aug. 09, 2023).

[8] D. D. Gu, W. Meiners, K. Wissenbach, and R. Poprawe, ‘Laser additive manufacturing of metallic components: materials, processes and mechanisms’, *International Materials Reviews*, vol. 57, no. 3, pp. 133–164, May 2012, doi: 10.1179/1743280411Y.0000000014.

[9] T. D. Ngo, A. Kashani, G. Imbalzano, K. T. Q. Nguyen, and D. Hui, ‘Additive manufacturing (3D printing): A review of materials, methods, applications and challenges’, *Composites Part B: Engineering*, vol. 143, pp. 172–196, Jun. 2018, doi: 10.1016/j.compositesb.2018.02.012.

[10] T. DebRoy *et al.*, ‘Additive manufacturing of metallic components – Process, structure and properties’, *Progress in Materials Science*, vol. 92, pp. 112–224, Mar. 2018, doi: 10.1016/j.pmatsci.2017.10.001.

[11] S. W. Williams, F. Martina, A. C. Addison, J. Ding, G. Pardal, and P. Colegrove, ‘Wire + Arc Additive Manufacturing’, *Materials Science and Technology*, vol. 32, no. 7, pp. 641–647, May 2016, doi: 10.1179/1743284715Y.0000000073.

[12] R. L. Truby and J. A. Lewis, ‘Printing soft matter in three dimensions’, *Nature*, vol. 540, no. 7633, pp. 371–378, Dec. 2016, doi: 10.1038/nature21003.

[13] S. M. Thompson, L. Bian, N. Shamsaei, and A. Yadollahi, ‘An overview of Direct Laser Deposition for additive manufacturing; Part I:

Transport phenomena, modeling and diagnostics', *Additive Manufacturing*, vol. 8, pp. 36–62, Oct. 2015, doi: 10.1016/j.addma.2015.07.001.

[14] A. J. K. Väisänen, M. Hyttinen, S. Ylönen, and L. Alonen, 'Occupational exposure to gaseous and particulate contaminants originating from additive manufacturing of liquid, powdered, and filament plastic materials and related post-processes', *Journal of Occupational and Environmental Hygiene*, vol. 16, no. 3, pp. 258–271, Mar. 2019, doi: 10.1080/15459624.2018.1557784.

[15] S. Sittichompoo, S. Kanagalingam, L. E. J. Thomas-Seale, A. Tsolakis, and J. M. Herreros, 'Characterization of particle emission from thermoplastic additive manufacturing', *Atmospheric Environment*, vol. 239, p. 117765, Oct. 2020, doi: 10.1016/j.atmosenv.2020.117765.

[16] R. Pernetti, F. Galbusera, A. Cattenone, E. Bergamaschi, B. Previtali, and E. Oddone, 'Characterizing Nanoparticle Release Patterns of Laser Powder Bed Fusion in Metal Additive Manufacturing: First Step Towards Mitigation Measures', *Annals of Work Exposures and Health*, vol. 67, no. 2, pp. 252–265, Feb. 2023, doi: 10.1093/annweh/wxac080.

[17] A. B. Stefaniak *et al.*, 'Large-Format Additive Manufacturing and Machining Using High-Melt-Temperature Polymers. Part I: Real-Time Particulate and Gas-Phase Emissions', *ACS Chem. Health Saf.*, vol. 28, no. 3, pp. 190–200, May 2021, doi: 10.1021/acs.chas.0c00128.

[18] R. Pernetti, S. Maffia, B. Previtali, and E. Oddone, 'Assessment of nanoparticle emission in additive manufacturing: Comparing wire and powder laser metal deposition processes', *Journal of Occupational and Environmental Hygiene*, pp. 1–7, Jun. 2023, doi: 10.1080/15459624.2023.2208649.

[19] N. Afshar-Mohajer, C.-Y. Wu, T. Ladun, D. A. Rajon, and Y. Huang, 'Characterization of particulate matters and total VOC emissions from a binder jetting 3D printer', *Building and Environment*, vol. 93, pp. 293–301, Nov. 2015, doi: 10.1016/j.buildenv.2015.07.013.

[20] M. Vogt, J. Schlichter, F. Aschersleben, T. Abraham, L. Wolf, and C. Herrmann, 'Integration of cyber-physical HVAC systems in Incremental Manufacturing to improve Energy Efficiency and Air Quality', *Procedia CIRP*, vol. 104, pp. 482–487, 2021, doi: 10.1016/j.procir.2021.11.081.

[21] B. Lin, B. Recke, J. K. H. Knudsen, and S. B. Jørgensen, 'A systematic approach for soft sensor development', *Computers & Chemical Engineering*, vol. 31, no. 5, pp. 419–425, May 2007, doi: 10.1016/j.compchemeng.2006.05.030.

[22] G. A. Montague, A. J. Morris, and M. T. Tham, 'Enhancing bioprocess operability with generic software sensors', *Journal of Biotechnology*, vol. 25, no. 1, pp. 183–201, Aug. 1992, doi: 10.1016/0168-1656(92)90114-O.

[23] G. C. Goodwin, 'Predicting the performance of soft sensors as a route to low cost automation', *Annual Reviews in Control*, vol. 24, pp. 55–66, Jan. 2000, doi: 10.1016/S1367-5788(00)90013-0.

[24] S. H. Hasnen, M. Shahid, H. Zabiri, and S. A. A. Taqvi, 'Semi-supervised adaptive PLS soft-sensor with PCA-based drift correction method for online valuation of NOx emission in industrial water-tube boiler', *Process Safety and Environmental Protection*, vol. 172, pp. 787–801, Apr. 2023, doi: 10.1016/j.psep.2023.02.080.

[25] R. Liu, Z. Rong, B. Jiang, Z. Pang, and C. Tang, 'Soft Sensor of 4-CBA Concentration Using Deep Belief Networks with Continuous Restricted Boltzmann Machine', in *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, Nov. 2018, pp. 421–424. doi: 10.1109/CCIS.2018.8691166.

[26] W. Fan, F. Si, S. Ren, C. Yu, Y. Cui, and P. Wang, 'Integration of continuous restricted Boltzmann machine and SVR in NOx emissions prediction of a tangential firing boiler', *Chemometrics and Intelligent Laboratory Systems*, vol. 195, p. 103870, Dec. 2019, doi: 10.1016/j.chemolab.2019.103870.

[27] P. Lian, H. Liu, X. Wang, and R. Guo, 'Soft sensor based on DBN-IPSO-SVR approach for rotor thermal deformation prediction of rotary air-preheater', *Measurement*, vol. 165,

p. 108109, Dec. 2020, doi: 10.1016/j.measurement.2020.108109.

[28] Q. Sun and Z. Ge, ‘A Survey on Deep Learning for Data-Driven Soft Sensors’, *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 5853–5866, Sep. 2021, doi: 10.1109/TII.2021.3053128.

[29] D. Martin, N. Kühl, and G. Satzger, ‘Virtual Sensors’, *Bus Inf Syst Eng*, vol. 63, no. 3, pp. 315–323, Jun. 2021, doi: 10.1007/s12599-021-00689-w.

[30] ‘Virtual Sensor Definition | IoT ONE Digital Transformation Advisors’, *IoT ONE*. https://www.iotone.com/term/virtual-sensor/t480 (accessed Aug. 10, 2023).

[31] S. Yoon, ‘Virtual sensing in intelligent buildings and digitalization’, *Automation in Construction*, vol. 143, p. 104578, Nov. 2022, doi: 10.1016/j.autcon.2022.104578.

[32] A. Dogeanu *et al.*, ‘Real time monitoring network demonstrator for air quality management’, in *2019 International Conference on ENERGY and ENVIRONMENT (CIEM)*, Timisoara, Romania: IEEE, Oct. 2019, pp. 459–463. doi: 10.1109/CIEM46456.2019.8937625.

[33] M. Leidinger, T. Sauerwald, W. Reimringer, G. Ventura, and A. Schütze, ‘Selective detection of hazardous VOCs for indoor air quality applications using a virtual gas sensor array’, *Journal of Sensors and Sensor Systems*, vol. 3, no. 2, pp. 253–263, Oct. 2014, doi: 10.5194/jsss-3-253-2014.

[34] A. Benabbas *et al.*, ‘Measure particulate matter by yourself: data-quality monitoring in a citizen science project’, *Journal of Sensors and Sensor Systems*, vol. 8, no. 2, pp. 317–328, Nov. 2019, doi: 10.5194/jsss-8-317-2019.

[35] H. Liu and C. Yoo, ‘A robust localized soft sensor for particulate matter modeling in Seoul metro systems’, *Journal of Hazardous Materials*, vol. 305, pp. 209–218, Mar. 2016, doi: 10.1016/j.jhazmat.2015.11.051.

[36] H. Liu, C. Yang, M. Huang, and C. Yoo, ‘Soft sensor modeling of industrial process data using kernel latent variables-based relevance vector machine’, *Applied Soft Computing*, vol. 90, p. 106149, May 2020, doi: 10.1016/j.asoc.2020.106149.

[37] J. Loy-Benitez, S. Tariq, H. T. Nguyen, U. Safder, K. Nam, and C. Yoo, ‘Neural circuit policies-based temporal flexible soft-sensor modeling of subway PM2.5 with applications on indoor air quality management’, *Building and Environment*, vol. 207, p. 108537, Jan. 2022, doi: 10.1016/j.buildenv.2021.108537.

[38] J. Loy-Benitez, S. Heo, and C. Yoo, ‘Soft sensor validation for monitoring and resilient control of sequential subway indoor air quality through memory-gated recurrent neural networks-based autoencoders’, *Control Engineering Practice*, vol. 97, p. 104330, Apr. 2020, doi: 10.1016/j.conengprac.2020.104330.

[39] K. P, P. Selvaraj, V. K. Burugari, C. P. K, and S. P. S, ‘An energy efficient approach of deep learning based soft sensor for air quality management’, *Measurement: Sensors*, vol. 24, p. 100460, Dec. 2022, doi: 10.1016/j.measen.2022.100460.

[40] M. A. Zaidan *et al.*, ‘Intelligent Calibration and Virtual Sensing for Integrated Low-Cost Air Quality Sensors’, *IEEE Sensors J.*, vol. 20, no. 22, pp. 13638–13652, Nov. 2020, doi: 10.1109/JSEN.2020.3010316.

[41] P. Kadlec, R. Grbić, and B. Gabrys, ‘Review of adaptation mechanisms for data-driven soft sensors’, *Computers & Chemical Engineering*, vol. 35, no. 1, pp. 1–24, Jan. 2011, doi: 10.1016/j.compchemeng.2010.07.034.

[42] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 1st ed. in Wiley Series in Probability and Statistics. Wiley, 2002. doi: 10.1002/9781119013563.

[43] A. Hyvärinen and E. Oja, ‘Independent component analysis: algorithms and applications’, *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, Jun. 2000, doi: 10.1016/S0893-6080(00)00026-5.

[44] I. T. Jolliffe and J. Cadima, ‘Principal component analysis: a review and recent developments’, *Phil. Trans. R. Soc. A.*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.

[45] J. F. Hair Jr, M. Sarstedt, L. Hopkins, and V. G. Kuppelwieser, ‘Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research’, *European Business Review*, vol. 26, no. 2, pp. 106–121, Mar. 2014, doi: 10.1108/EBR-10-2013-0128.

[46] J. F. Hair, C. M. Ringle, and M. Sarstedt, 'PLS-SEM: Indeed a Silver Bullet', *Journal of Marketing Theory and Practice*, vol. 19, no. 2, pp. 139–152, Apr. 2011, doi: 10.2753/MTP1069-6679190202.

[47] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, 'A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking', *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002, doi: 10.1109/78.978374.

[48] S. J. Julier and J. K. Uhlmann, 'Unscented Filtering and Nonlinear Estimation', *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004, doi: 10.1109/JPROC.2003.823141.

[49] E. P. Carden and P. Fanning, 'Vibration Based Condition Monitoring: A Review', *Structural Health Monitoring*, vol. 3, no. 4, pp. 355–377, Dec. 2004, doi: 10.1177/1475921704047500.

[50] Wei Fan and Pizhong Qiao, 'Vibration-based Damage Identification Methods: A Review and Comparative Study', *Structural Health Monitoring*, vol. 10, no. 1, pp. 83–111, Jan. 2011, doi: 10.1177/1475921710365419.

[51] R. Haber and H. Unbehauen, 'Structure identification of nonlinear dynamic systems—A survey on input/output approaches', *Automatica*, vol. 26, no. 4, pp. 651–677, Jul. 1990, doi: 10.1016/0005-1098(90)90044-I.

[52] B. Walczak and D. L. Massart, 'Dealing with missing data', *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 1, pp. 15–27, Sep. 2001, doi: 10.1016/S0169-7439(01)00131-9.

[53] T. Lorido-Botran, J. Miguel-Alonso, and J. A. Lozano, 'A Review of Auto-scaling Techniques for Elastic Applications in Cloud Environments', *J Grid Computing*, vol. 12, no. 4, pp. 559–592, Dec. 2014, doi: 10.1007/s10723-014-9314-7.

[54] E. Gémez, H. Unbehauen, P. Kortmann, and S. Peters, 'Fault Detection and Diagnosis with the Help of Fuzzy-Logic and with Application to a Laboratory Turbogenerator', *IFAC Proceedings Volumes*, vol. 29, no. 1, pp. 6554–6559, Jun. 1996, doi: 10.1016/S1474-6670(17)58734-0.

[55] S. G. Carmer and M. R. Swanson, 'An Evaluation of Ten Pairwise Multiple Comparison Procedures by Monte Carlo Methods', *Journal of the American Statistical Association*, vol. 68, no. 341, pp. 66–74, Mar. 1973, doi: 10.1080/01621459.1973.10481335.

[56] G. Widmer and M. Kubat, 'Learning in the presence of Concept drift and hidden context', *Machine Learning*, vol. 23, no. 1, pp. 69–101, 1996, doi: 10.1023/A:1018046501280.

[57] K. Warne, G. Prasad, S. Rezvani, and L. Maguire, 'Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion', *Engineering Applications of Artificial Intelligence*, vol. 17, no. 8, pp. 871–885, Dec. 2004, doi: 10.1016/j.engappai.2004.08.020.

[58] 'An introduction to variable and feature selection | The Journal of Machine Learning Research'. https://dl.acm.org/doi/10.5555/944919.944968 (accessed Aug. 11, 2023).

[59] B. Gabrys, 'Neuro-fuzzy approach to processing inputs with missing values in pattern recognition problems', *International Journal of Approximate Reasoning*, vol. 30, no. 3, pp. 149–179, Sep. 2002, doi: 10.1016/S0888-613X(02)00070-1.

[60] N. R. Pal and S. K. Pal, 'A review on image segmentation techniques', *Pattern Recognition*, vol. 26, no. 9, pp. 1277–1294, Sep. 1993, doi: 10.1016/0031-3203(93)90135-J.

[61] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, 'On Clustering Validation Techniques', *Journal of Intelligent Information Systems*, vol. 17, no. 2/3, pp. 107–145, 2001, doi: 10.1023/A:1012801612483.

[62] C. J. C. Burges, 'A tutorial on support vector machines for pattern recognition', *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.

[63] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[64] J. Schmidhuber, 'Deep learning in neural networks: An overview', *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.

[65] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, ‘Generative Adversarial Networks: An Overview’, *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.

[66] J. Elith, J. R. Leathwick, and T. Hastie, ‘A working guide to boosted regression trees’, *J Anim Ecology*, vol. 77, no. 4, pp. 802–813, Jul. 2008, doi: 10.1111/j.1365-2656.2008.01390.x.

[67] R. Caruana, ‘Multitask Learning’, *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997, doi: 10.1023/A:1007379606734.

[68] M. Belgiu and L. Drăguţ, ‘Random forest in remote sensing: A review of applications and future directions’, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

[69] L. B. Booker, D. E. Goldberg, and J. H. Holland, ‘Classifier systems and genetic algorithms’, *Artificial Intelligence*, vol. 40, no. 1–3, pp. 235–282, Sep. 1989, doi: 10.1016/0004-3702(89)90050-7.

[70] S. S. Rao, *Engineering Optimization: Theory and Practice*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2009. doi: 10.1002/9780470549124.

[71] R. Xu and D. WunschII, ‘Survey of Clustering Algorithms’, *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005, doi: 10.1109/TNN.2005.845141.

[72] T. Kohonen, ‘The self-organizing map’, *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990, doi: 10.1109/5.58325.

[73] ‘sklearn.linear_model.Lasso’, *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.linear_model.Lasso.html (accessed Aug. 10, 2023).

[74] ‘sklearn.linear_model.BayesianRidge’, *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.linear_model.BayesianRidge.html (accessed Aug. 10, 2023).

[75] ‘sklearn.linear_model.ElasticNet’, *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.linear_model.ElasticNet.html (accessed Aug. 10, 2023).

[76] ‘9. Model persistence’, *scikit-learn*. https://scikit-learn/stable/model_persistence.html (accessed Aug. 10, 2023).