

Article

Investigation of a Sparse Autoencoder-Based Feature Transfer Learning Framework for Hydrogen Monitoring Using Microfluidic Olfaction Detectors

Hamed Mirzaei ^{1,†} , Milad Ramezankhani ^{1,†}, Emily Earl ², Nishat Tasnim ², Abbas S. Milani ^{1,*} 
and Mina Hoorfar ^{2,*} 

¹ School of Engineering, University of British Columbia Okanagan Campus, Kelowna, BC V1V 1V7, Canada

² Department of Mechanical Engineering, University of Victoria, Victoria, BC V8P 5C2, Canada

* Correspondence: abbas.milani@ubc.ca (A.S.M.); mhoorfar@uvic.ca (M.H.)

† These authors contributed equally to this work.

Abstract: Alternative fuel sources, such as hydrogen-enriched natural gas (HENG), are highly sought after by governments globally for lowering carbon emissions. Consequently, the recognition of hydrogen as a valuable zero-emission energy carrier has increased, resulting in many countries attempting to enrich natural gas with hydrogen; however, there are rising concerns over the safe use, storage, and transport of H₂ due to its characteristics such as flammability, combustion, and explosivity at low concentrations (4 vol%), requiring highly sensitive and selective sensors for safety monitoring. Microfluidic-based metal–oxide–semiconducting (MOS) gas sensors are strong tools for detecting lower levels of natural gas elements; however, their working mechanism results in a lack of real-time analysis techniques to identify the exact concentration of the present gases. Current advanced machine learning models, such as deep learning, require large datasets for training. Moreover, such models perform poorly in data distribution shifts such as instrumental variation. To address this problem, we proposed a Sparse Autoencoder-based Transfer Learning (SAE-TL) framework for estimating the hydrogen gas concentration in HENG mixtures using limited datasets from a 3D printed microfluidic detector coupled with two commercial MOS sensors. Our framework detects concentrations of simulated HENG based on time-series data collected from a cost-effective microfluidic-based detector. This modular gas detector houses metal–oxide–semiconducting (MOS) gas sensors in a microchannel with coated walls, which provides selectivity based on the diffusion pace of different gases. We achieve a dominant performance with the SAE-TL framework compared to typical ML models (94% R-squared). The framework is implementable in real-world applications for fast adaptation of the predictive models to new types of MOS sensor responses.

Keywords: hydrogen detection; HENG; transfer learning; sparse autoencoder; microfluidic gas sensor



Citation: Mirzaei, H.; Ramezankhani, M.; Earl, E.; Tasnim, N.; Milani, A.S.; Hoorfar, M. Investigation of a Sparse Autoencoder-Based Feature Transfer Learning Framework for Hydrogen Monitoring Using Microfluidic Olfaction Detectors. *Sensors* **2022**, *22*, 7696. <https://doi.org/10.3390/s22207696>

Academic Editor: James Covington

Received: 15 August 2022

Accepted: 6 October 2022

Published: 11 October 2022

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



Copyright: © 2022 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

Blending hydrogen into the existing natural gas pipeline network has to reduce greenhouse gas emissions from natural gas production, distribution, and consumption fields [1]. Reports show that conventional NG pipelines can carry a blend of natural gas and hydrogen of up to approximately 15% by volume of hydrogen while only requiring modest modifications to the pipeline [2]. While hydrogen is a valuable and clean alternative for carbon-based fuels, its unique physicochemical properties make it a highly permeable and explosive gas that requires precise monitoring. Hydrogen gas is odorless, colorless, and non-toxic. Therefore, it cannot be sensed by the human olfactory system. Due to this, enhanced safety measures must be put in place to monitor its presence and concentration in various environments. These sensors are required to monitor leakage of hydrogen, and measure hydrogen concentration in all processes involving production, transportation, and storage. The gold standard methods used to analyze gaseous molecules are gas

chromatography-based and mass spectrometry-based technologies [3], but these platforms are relatively large, expensive, and somewhat slow.

On the other hand, hydrogen sensors are cost-effective, small, easy to operate, and can be highly sensitive and selective. Hydrogen sensors have been produced for several decades. While several commercialized hydrogen sensors are already available, the efforts to improve them for continuous and precise real-time monitoring for the future hydrogen-based economy are ongoing. These efforts focus on improving the signal's duration, the sensors' selectivity and sensitivity, and the miniaturization of the sensors.

Numerous technologies have been developed to detect hydrogen, such as optical sensors [4,5], electrochemical sensor fields [6,7], catalytic sensors [8,9], work function sensors [10,11], and resistance-based sensors [12,13]. Among these devices, resistance-based sensors, specifically metal-oxide-semiconductor (MOS) sensors, have gained significant attention in gas detection due to their relatively small size, low cost, ease of operation, and high sensitivity [14]. As the sensing mechanism in a MOS sensor is based on the adsorption and the consumption of the oxygen molecules (by reducing gas in *p*-type semiconductors and vice versa in *n*-type semiconductors) on the surface of the sensing layer, these sensors can detect a wide range of gaseous elements such as volatile organic compounds, hydrocarbons, and combustible gases, including hydrogen. However, this wide detection range limits specificity and sensitivity for certain analytes. To overcome this issue, a 3D-printed microfluidic channel can be used to delay the diffusion of the target gas onto the surface of the sensing layer. Using novel coating compositions and surface treatments on the channels' inner walls, MOS-microchannel platforms can noticeably enhance the selectivity of MOS sensors to various target gases [15].

The resulting response curves generated from MOS-based microfluidic detectors must be processed using statistical analysis, namely feature extraction, to more accurately estimate the hydrogen concentration. Many feature extraction methods implemented in the literature attempt to generate a condensed representation (i.e., dimensionality reduction) of the signals produced by sensors [16]. Given the limited available gas sensor data, the high-dimensional input feature space of the response curves must be translated into a low-dimensional feature representation to prevent the "curse of dimensionality" causing poor generalization performance (i.e., overfitting) in the Machine Learning (ML) model [17]. In one common approach, the desired features are extracted by the expert user based on domain knowledge, experience, and the geometric characteristics of the response curve [18]. The most commonly extracted features include the maximum response value, the area under the curve, and the sensor response's rising and falling slopes/time [17]. The drawback of this method is that the feature extraction process needs to be performed manually, as the meaningful features for each gas/sensor response curve can vary. It is also likely that some informative features may be overlooked and excluded from data analysis.

Unsupervised learning methods are powerful dimensionality reduction techniques used to learn and extract meaningful information from sensor responses [18]. Principal component instance and principal component analysis (PCA) are considered liable unsupervised techniques, widely used for reducing the dimensionality of the input feature space before being fed into the ML model. The principal components' direction in PCA is calculated to maximize the described variance of the original dataset [19]. As such, PCA is a popular choice for achieving low-level representations of data. Yin and Tian [20] utilized PCA to reduce the dimensions of gas sensor data for classifying Chinese drinks. The advantage of PCA over the above feature extraction methods is that it automatically learns the most important elements in the dataset while being robust to the presence of correlated input variables. However, PCA fails to extract useful features in highly non-linear feature spaces as it is limited to linear projections [17].

Similarly, other traditional dimensionality reduction approaches, such as manifold learning methods, fall short in learning high-level abstractions [21]. Deep learning models offer more flexibility than conventional ML models and non-linear data structures. In other words, due to the non-linearity in their activation functions, they are capable of learning

complex operations and highly non-linear feature representations. Autoencoders (AEs) are deep unsupervised learning methods designed for learning dense latent features and dimensionality reduction [22].

Once the latent features have been determined, the pre-processed data usually is fed to a classification/regression model. In gas detection, efforts have been made to improve the performance of electronic nose (e-nose) devices by incorporating deep learning (DL) models. Zhao et al. [23] utilized a deep learning model to differentiate various Chinese liqueurs based on the responses from electronic noses. Ma et al. [24] used the images produced by a sensor array to train a Convolutional Neural Network (CNN) to detect unknown gases. The result indicates that CNN's performance is superior to traditional ML approaches. Despite their outstanding performance, deep learning models require large datasets to be trained properly. In the absence of a sufficient amount of data, deep learning models often overfit the training dataset (due to many trainable parameters), thus generalizing poorly on new unseen instances.

One way to mitigate the performance-declining effect of limited data is to implement Transfer Learning (TL) [25]. TL leverages knowledge from related domains and incorporates it into the learning process for tasks of interest with limited data. This significantly reduces the dependency complex models, such as DL-based methods, have on large datasets while improving the model's generalization performance [25]. Yan and Zhang [26] developed a sample-based transfer learning using Ridge and Logistic regressions for tackling the domain distribution shifts caused by sensor drifts and instrumental variation. To address the sensor drift, the proposed framework by Yi et al. [27] initially minimizes the domains' distribution inconsistency by minimizing the Maximum Mean Discrepancy (MMD), followed by an adaptive Extreme Machine Learning (EML) classifier.

Although some research has been performed on mitigating the effect of instrumental variation and sensor drift on ML performance for various types of gases and e-noses, investigating and addressing such a phenomenon for hydrogen-enriched natural gas (HENG) mixtures remains unexplored. In this paper, a Sparse Autoencoder-based Transfer Learning (SAE-TL) is developed to address the issue of the sensor response shift in an in-house 3D printed microfluidic channel coupled with two commercial MOS sensors for estimating the hydrogen gas concentration in HENG mixtures using limited datasets. Using source data, the framework first learns a low-level representation of the response curves by SAE through an unsupervised learning procedure. Then, the learned words are used in the TL portion of the framework to handle the domain shifts and learn a reliable regressor for a new sensor model with very few available data.

2. SAE-TL Framework

The schematic of the proposed framework is illustrated in Figure 1, and the details are elaborated in the following sub-sections.

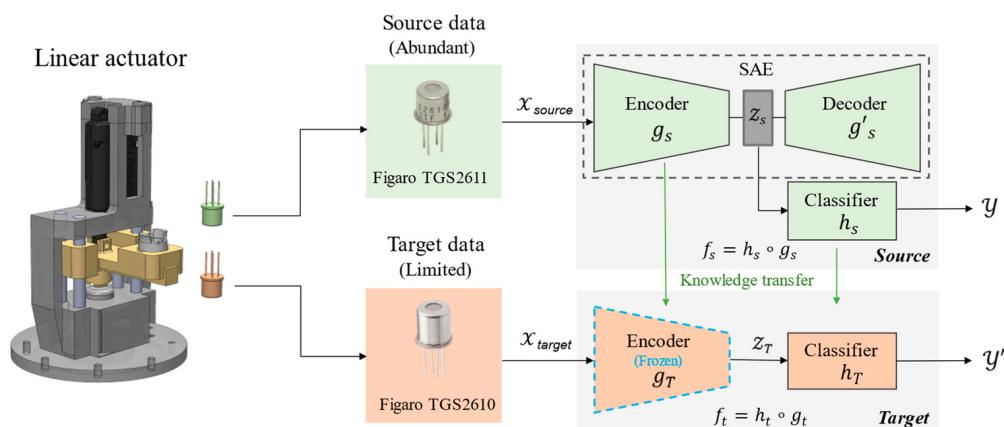


Figure 1. Schematic of the proposed SAE-TL framework.

2.1. Sparse Autoencoder (SAE)

An AE is comprised of an encoder and a decoder. The encoder, which consists of the input and hidden layers, receives high-dimensional raw data and transforms it into a low-level latent representation, while the decoder (the hidden and output layers) is responsible for reconstructing the original data from the hidden layer output, i.e., condensed representation. The hidden layer typically contains fewer neurons than the input and output layers, so a denser representation of the raw data can be obtained [18]. AEs learn to minimize the distance between the raw signal x and its reconstructed version $\hat{x}$ by minimizing the reconstruction loss,

$$J_{AE}(\theta) = \sum_{i=1}^n L(x_i, \hat{x}_i) \quad (1)$$

where $\theta = \{W, b, W', b'\}$ denotes the weights and biases of the encoder and decoder, respectively, and n is the dataset size.

Due to the fact that AEs are designed to identify meaningful features from raw data and reconstruct the original data at the output layer, they are prone to copying the raw data from the input layer to the output layer without extracting any useful features [22]. To avoid this, one approach is to monitor the activation of each neuron in the hidden layer and penalize the units that have an activation higher than the specified threshold using a sparsity loss term. As a result, the SAE learns useful representations while attempting to avoid large activations in the hidden layer. Kullback–Leibler (KL) divergence is used as the sparsity loss and can be represented as (2).

$$KL(q||p_k) = q \log \frac{q}{p_k} - (1 - q) \log \frac{1 - q}{1 - p_k} \quad (2)$$

It measures the distance between the predefined sparse parameter q (usually set to be a small value) and the average activation of each hidden unit:

$$p_k = \frac{1}{n} \sum_{i=1}^n [s_f(b_k + W_k x_i)] \quad (3)$$

where s_f denotes the hidden layers activation function. Thus, the cost function of the SAE can be written by combining the AE cost function and the KL divergence measure as follows:

$$J_{SAE}(\theta) = \sum_{i=1}^n L(x_i, \hat{x}_i) + KL(q||p_k) \quad (4)$$

2.2. Transfer Learning

Deep learning approaches have been shown to be a powerful tool in deriving important features from the gas sensor and e-nose response curves [28,29]. Despite the promising results, deep learning approaches require large amounts of data for training and are usually ill-equipped to handle slight shifts in data distribution. For example, implementing a new sensor with similar but different specifications will cause the model to be unable to perform accurate prediction tasks. One solution to alleviate this dependency on large datasets is to utilize TL [28]. In a TL framework, the knowledge gained from a model that is trained on sufficient data (the source) is used for improving the performance of a model that is trained on the task of interest that has limited data (the target) [30]. When considering HENG mixtures detection tasks, historical experimental data has been generated from a specific type of sensor. An ML model trained on such data may be ineffective when used against a new sensor response (i.e., domain shift), as it will exhibit different behavior. However, it is also known that the response curves of gas sensors share considerable similarities, suggesting that a model trained on historical data can still be useful as it conveys the learned knowledge about the general behavior of the response curve. This knowledge can significantly decrease the data required for learning low-level available

feature representations if transferred to the target model via TL. In Neural Networks (NN), TL can be incorporated by transferring the fine-tuned weights from the source model to the target model. Using this approach, the target network requires only a few gradient steps using the limited available data to reach optimal performance [31].

For an input space X and a label space Y , the ML objective predictive function $f: X \rightarrow Y$ can be broken down into two corresponding functions: $f = h \circ g$. While $limitg: X \rightarrow Z$ maps the input space to a latent low-dimensional feature space, Z , the embedding function $limith: Z \rightarrow Y$ is a regressor (a predictive function) that predicts the output y using the feature space. It has been shown that, when training NNs on different but related datasets, the first layers of the networks learn similar low-level feature representations. This general behavior is seen in all networks regardless of the model specifications (e.g., the cost function and the input data structure) [32]. On the contrary, the final layers of networks developed using related data sets exhibit completely different behavior. Towards their final layers, each network becomes more specific to the task they are being trained on. This phenomenon is observed in networks trained on various types of data (e.g., tabular, image, and text) [33].

For an SAE to learn low-level feature representations of gas sensors' response curves, it requires training the SAE using a large dataset. In cases where data is limited, i.e., a new sensor is implemented (instrumental variation), training the SAE and the classifier is not immediately possible. One infeasible remedy is to generate a large dataset using the new sensor, which causes high temporal and financial costs. A more efficient solution is to implement TL. In the proposed TL framework, low-level feature representations of the source sensor's response curves are initially learned using SAE.

Once learned, the SAE's encoder (g_s) is connected to a randomly initialized multilayer perceptron (MLP) (regressor, h_s) to form the source model. In a supervised regime, the labeled source data is used to train the source model in two steps. The initial weights of the target regressor produce large error gradients during the back-propagation step, which can destructively modify the learned weights of the transferred encoder. Initially, the encoder's weights are fixed (frozen) to avoid large error gradients. After a few epochs, as the regressor's weights are stable and the error gradient shrinks, the whole model becomes unfrozen and is trained until it reaches its optimal state. Once the source network is developed ($f_s = h_s \circ g_s$), it is transferred to the target model (Figure 1). Next, the target model needs to be fine-tuned using the limited available target data. This step is necessary for the model to learn more specific features/trends of the target task, boosting its overall performance. However, training the whole network requires a large dataset. Therefore, the target encoder (h_T), transferred from the source model, is frozen during the fine-tuning process. This ensures that the knowledge learned from the general layers of the source model is transferred and remains intact (i.e., avoiding forgetting the learned knowledge). The target regressor (h_T), which contains more task-specific layers, will be fine-tuned with the target data. This allows the SAE to bypass the data-intensive procedure of learning low-level features, resulting in a network that requires only a small dataset to be re-calibrated to the desired task [34].

3. Experimental Verification

3.1. Case Study: Hydrogen Gas Detection Using a Microfluidic Detector

Hydrogen is a light and odorless gas; it is challenging to determine the concentration in the mixture of HENG in real time using commercial sensors [35]. The use of a microfluidic channel for the detection of gases based on adsorption/desorption phenomena provides the apparatus with enough delay in the diffusion time and, consequently, in the response time to differentiate between various gases [15,28,36]. The schematic diagram of the experimental sensing apparatus is illustrated in Figure 2. The sensor platform consists of a MOS sensor (Figaro TGS2610 or Figaro TGS2611), explosion-proof valves, a mixing chamber, mass flow controllers (MFCs), an automated sensor housing carrier, electronic components, and tubing. By calculating the chamber size and the passing flow rate, this

automatic setup supplies the mixing chamber with the desired concentration of a single gas or mixture of target analytes. Exhaust valves have been used to regulate residual gas pressure, which will be trapped between the MFCs and the chamber inlet valves. Before being exposed to the sensor, the gas mixture rests in the mixing chamber to ensure a homogenous state. During this period, the data collection process begins to generate a baseline signal for the sensor. After a homogenous state is attained, the microchannel inlet is exposed to the gas chamber by opening a magnetic valve (demonstrated in Figure 3). The gas mixture is then diffused through the microfluidic channel and reaches the sensing layer. Unique raw data for each experiment will be automatically collected in a RaspberryPi 4 and progressed in our TL framework. Due to the specific diffusion times of different gases and the adsorption/desorption phenomena that occur on the channel walls, individual responses are obtained for every gas mixture [15], providing the sensing platform with improved selectivity. Because of a large number of tests and to avoid a different MOS sensor drift impact [37], Figaro TGS2610 and TGS2611 sensors were switched after every five experiments. In order to not only make changing sensors feasible but also to ensure that the microchannel effect on gas responses was kept constant, a modular 3D-printed housing system was designed (Figure 4), allowing for different sensors to be mounted at the end of the microfluidic channel.

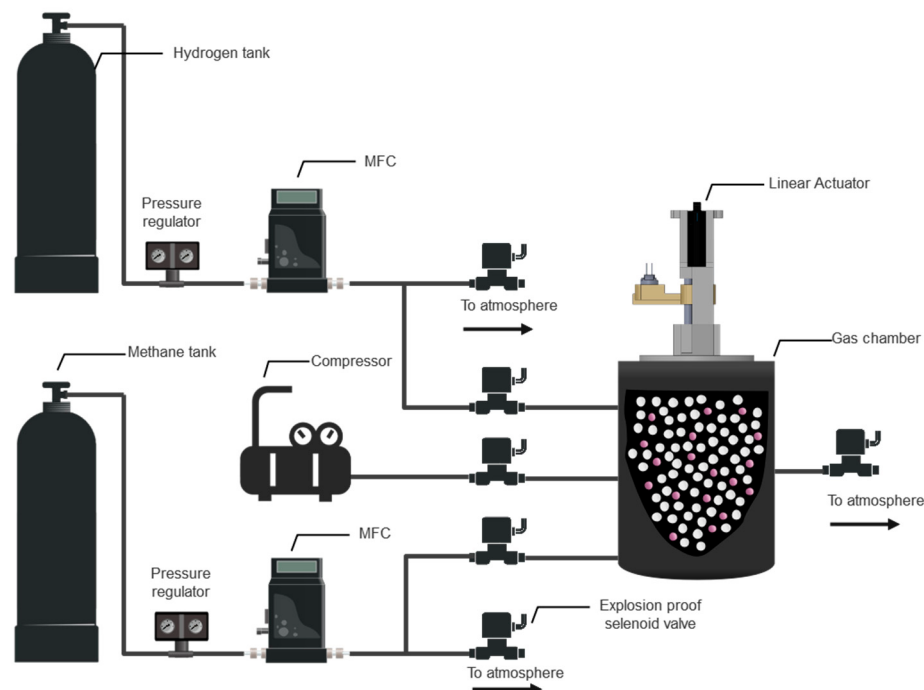


Figure 2. The schematic diagram of the automated experimental setup, designed and built for sample generation and data collection.

Based on a previous study to maximize the selectivity between VOCs, the channel length was set to 30 mm, with a width of 3 mm and a height of 1 mm. As reported in previous and parallel works done by our research group, these microchannel-MOS gas detectors can provide specific patterns for various gases and are used in several gas detection applications, such as wine identification, natural gas detection, hydrogen sulfide monitoring in sewer systems, and differentiating between VOCs [15,28,36,38].

3.2. Sensor Characteristics

The utilized sensors, TGS-2610 and TGS-2611, are semiconductor-based, and their output is raw voltages from the sensing layer before, during, and after exposure to a target gas, as illustrated in Figure 3. Conventional HENG monitoring methods have proven to have a 90–270 s response time (including the recovery time) [39–43]. In this work,

taking advantage of a microfluidic channel working as a delaying agent, exponential shape responses are generated from the interaction between the hydrogen-enriched natural gas and semiconductor-based and commercially available gas sensors. As demonstrated in Table 1, the response time for this platform is as low as 150 s. This relatively short and real-time response ensures a fast data collection procedure as part of the training of the proposed learning framework. Table 1 summarizes and compares key characteristics of available industrial and pilot HENG monitoring methods.

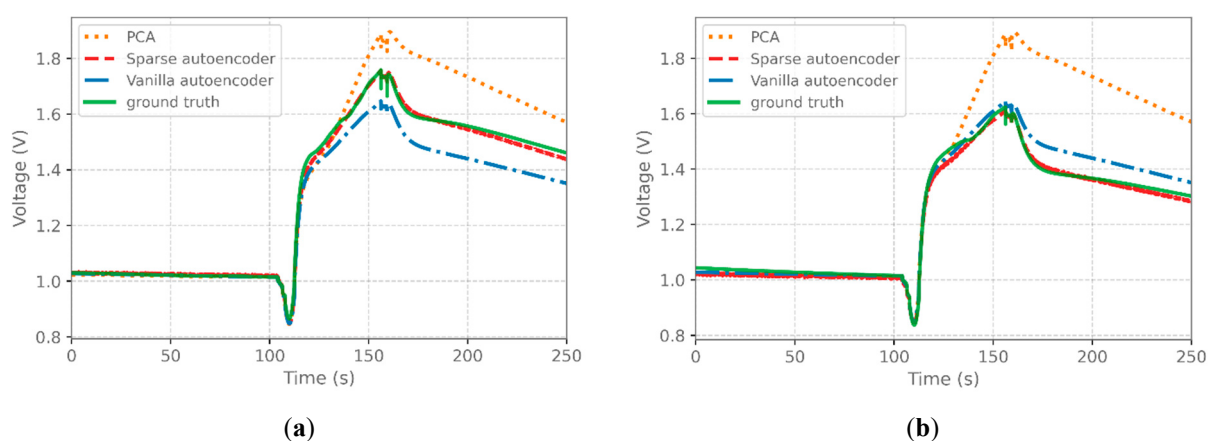


Figure 3. Comparison of reconstruction accuracy of SAE, AE, and PCA for gas mixture samples with (a) 810 (90) ppm and (b) 90 (10) ppm Hydrogen (Methane).

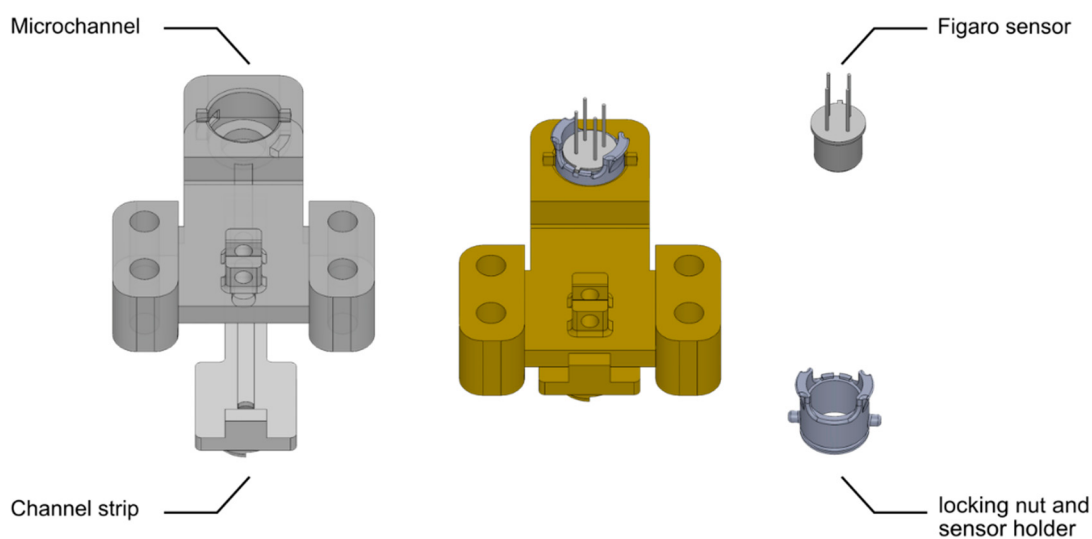


Figure 4. Parts and assembly of the 3D printed microchannel and the modular housing.

Table 1. Comparison of select hydrogen and natural gas mixture monitoring platforms by key characteristics.

Ref	Sensor Type	Limit of Detection	Power Consumption	Time from Collection to Results
[39]	Carbon nanotube/SiO ₂	Not reported	Not reported	1132 s
[40]	Calorimetric Pd/ θ -Al ₂ O ₃	200 ppm	0.12 W	Not reported
[41]	Au/SnO ₂ , Pt/Cu/SnO ₂	500 ppm	200 mW	180 s
[42]	Pd/Au optical sensor	987 ppm	Not reported	90 s
[43]	Semiconductor, Catalytic, electrochemical	200 ppm	Not reported	190–270 s
This Work	Semiconductor	89 ppm	200 mW	150 s

3.3. Gas Mixture

Binary mixtures of methane and hydrogen were used for this study, as they are the two main components of HENG. The desired concentration of the HENG mixture was mixed from compressed gas tanks of 99.9% methane and hydrogen, respectively (Praxair Canada, Mississauga, ON, Canada). Due to safety considerations and reported optimal percentages of hydrogen in HENG, HENG mixtures of up to 10% (*v/v*) hydrogen were prepared. As illustrated in Figure 4, hydrogen and methane pressures were regulated to MFC's operating pressure to obtain a homogenous mixture. A specific amount of each gas was added to the gas chamber. Various methane and hydrogen mixture concentrations (20–1000 ppm) were obtained using the automated experimental setup to generate data for validating the proposed framework. In particular, 120 data points were generated from the source sensor (TGS2611), while 102 samples were collected from the target sensor (TGS2610). Each experiment was repeated three times to ensure the repeatability of the obtained responses.

3.4. SAE-TL Experimental Design

In SAE, symmetrical architecture concerning the latent representation layer (central hidden layer) is used, where the decoder mirrors the number of layers and neurons of the encoder. The main hidden layer contains 20 neurons, yielding a 20-dimensional representation of the raw data. The top-performing SAE encoder (and decoder) consists of two hidden layers with [500, 200] neurons. All coatings (except the decoder's output layer) are equipped with the RELU activation function. Mean Squared Error (MSE) is chosen as the loss function (SAE reconstruction loss), and an Adam optimizer with a learning rate of 0.001 is implemented. Early stopping with the patience of 50 is used as the regularization term. The sparsity threshold of $10e-5$ is selected for the latent representation layer. The TL portion of the framework contains two identical NNs, namely, the source and the target. Each network has an encoder with the same architecture as the SAE's encoder. The network's regressor is a four-layer NN with [10, 8, 6, 1] neurons. The RELU activation function and Adam optimizer with a learning rate of 0.001 were used for both networks. The whole framework is developed and trained using the Keras library in Python. For preparing the source network, the source data set is divided into training, validation, and testing (70%, 10%, 20%). For the target network, however, to mimic the available limited data in real-world scenarios, only 20% of data is allocated for training. The rest is kept to evaluate the generalization performance.

4. Results and Discussion

4.1. SAE Performance Evaluation

The performance of the SAE on the reconstruction of the sensor responses is compared with regular (vanilla) AE and PCA, and the results are summarized in Figure 3. For PCA, the first five principal components are used to reconstruct the original response. For SAE and AE, the decoder was used to map the low-level feature space to the actual distance (see Figure 1). The reconstructed responses of the SAE closely follow the ground truth curves. At the same time, both AE and PCA fall short in capturing the general behavior of the sensor response. The dominant performance of the SAE can be explained by its capability to capture non-linear representations (unlike PCA) and the effect of sparsity (regularization), which makes the model more robust against overfitting than AE.

The effect of the number of hidden layers (i.e., encoder layers + latent representation layer + decoder layers) of the AE and SAE on their reconstruction performance is also studied. Adding more hidden layers can give the model more flexibility to learn nonlinearity in the feature space better. On the other hand, it increases the number of hyperparameters, which may lead to overfitting during the training. Table 2 summarizes the model MSE loss when trained with 1–9 hidden layers. Though having a minor effect, increasing the number of hidden layers seems to reduce the reconstruction MSE reaching its minimum at five hidden layers. Adding more layers does not further reduce the models' loss, which

can indicate overfitting due to the large number of parameters added by each layer. For the rest of the paper, results of SAE and AE with five hidden layers are reported.

Table 2. Effect of number of hidden layers on the performance (MSE) of SAE and AE.

	Number of Hidden Layers				
	1	3	5	7	9
SAE	0.00035 ± 0.00001	0.00031 ± 0.00002	0.00024 ± 0.00001	0.00034 ± 0.00002	0.00039 ± 0.00001
AE	0.0017 ± 0.00002	0.001 ± 0.00001	0.0007 ± 0.00001	0.008 ± 0.0001	0.0008 ± 0.0001

4.2. TL Performance Evaluation

To evaluate the performance of the proposed TL method, it is compared with conventional state-of-the-art ML approaches. For the dimensionality reduction, PCA was performed on two datasets: (1) the limited target data and (2) the source and target data combined. The reduced representations, i.e., top five principal components, were then passed to a regression model, namely, a Support Vector Machine (SVM) [44], a Random Forest (RF) [45], and an XGBoost [46]. Mean Absolute Error (MAE) and R-squared of each model against the target test data are measured and summarized in Table 3.

Table 3. Generalization performance of the proposed TL method on test sensor data.

Encoder	Regressor	MAE (ppm)	R-Squared
SAE-TL	MLP	89.24	0.94
AE-TL	MLP	99.74	0.89
PCA (Source)	XGBoost	172.77	0.82
	RF	206.53	0.67
	SVM	202.47	0.63
PCA	XGBoost	172.77	0.82
(Source + Target)	RF	176.82	0.73
	SVM	109.88	0.87
	NN	121.25	0.84

SAE-TL demonstrates a clear dominant performance in estimating the hydrogen concentration with an MAE of 89.24. Combining the source and target datasets for the PCA input slightly improved the conventional models' performance. This is expected, as PCA, an unsupervised learning model, can benefit from a larger sample size to output a more robust representation. The AE-TL outperforms the PCA-based models, though it falls behind the SAE-TL due to its higher reconstruction loss (i.e., generating less-accurate representations). Finally, the NN directly trained on raw data (i.e., no dimensionality reduction) exhibits similar performance to that of the PCA-based models. The network is prone to overfitting due to the limited dataset, which reduces the model's generalization performance.

5. Conclusions

This paper presents a novel TL framework for predicting the HENG mixture concentration with limited data extracted from a microfluidic-based gas detector. Our in-house 3D-printed microfluidic channel works as a delay agent and provides a unique fingerprint response [15]. The proposed framework consists of an SAE for extracting useful low-dimensional representations from the high-dimensional raw detector response data, and a TL method for transferring the learned knowledge from historical data towards a model based on data from a new detector. The response curve data from two sensors have been collected to evaluate the proposed SAE-TL method's performance. The results clearly outline the proposed method's dominant performance compared with conventional ML models. Namely, the SAE outperformed AE and PCA in learning more useful features for dimensionality reduction. The implementation of TL drastically reduced the need for large amounts of data for training deep neural networks. The proposed learning framework can

effectively work with very limited data to learn the mapping between the sensor response and the gas concentration via the TL module in the framework. Conventional learning models require large datasets (especially since the input space, namely, sensor voltage response, is high-dimensional) to be fully trained and yield high accuracy. In contrast, our framework can enable a fast yet much less data-intensive training procedure to produce high predictive performance. In addition, many of the previous works have implemented conventional (and linear) dimensionality reduction methods such as PCA as part of the learning process; though effective, we have shown that by using a nonlinear counterpart (i.e. sparse autoencoder), a more informative low-dimensional representation of the input space can be achieved, directly leading to better predictive accuracy. Future studies are required to further validate the proposed framework's generalizability, and response data from other types of sensors can be collected and examined. Moreover, probabilistic models such as Gaussian processes [47] or Bayesian deep learning [48] can be utilized to quantify the model uncertainty.

Author Contributions: Conceptualization, H.M., M.R., M.H. and A.S.M.; methodology, H.M. and M.R.; software, E.E.; validation, H.M. and M.R.; formal analysis, H.M. and M.R.; investigation, H.M.; resources, H.M. and M.H.; data curation, M.R.; writing—original draft preparation, H.M. and M.R.; writing—review and editing, E.E., N.T., M.H. and A.S.M.; visualization, H.M.; supervision, M.H. and A.S.M.; project administration, H.M.; funding acquisition, M.H. and A.S.M. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sagar, S.M.V.; Agarwal, A.K. Hydrogen-Enriched Compressed Natural Gas: An Alternate Fuel for IC Engines. In *Advances in Internal Combustion Engine Research*; Springer: Singapore, 2018; pp. 111–134.
2. Melaina, M.W.; Antonia, O.; Penev, M. *Blending Hydrogen into Natural Gas Pipeline Networks: A Review of Key Issues*; NREL: Golden, CO, USA, 2013.
3. Sparkman, O.D.; Penton, Z.; Kitson, F.G. *Gas Chromatography and Mass Spectrometry: A Practical Guide*; Elsevier Inc.: Amsterdam, The Netherlands, 2011.
4. She, X.; Shen, Y.; Wang, J.; Jin, C. Pd films on soft substrates: A visual, high-contrast and low-cost optical hydrogen sensor. *Light Sci. Appl.* **2019**, *8*, 2047–7538. [[CrossRef](#)]
5. Aray, A.; Ranjbar, M.; Shokoufi, N.; Morshedi, A. Plasmonic fiber optic hydrogen sensor using oxygen defects in nanostructured molybdenum trioxide film. *Opt. Lett.* **2019**, *44*, 4773. [[CrossRef](#)]
6. Kalyakin, A.; Volkov, A.; Lyagaeva, J.; Medvedev, D.; Demin, A.; Tsiakaras, P. Combined amperometric and potentiometric hydrogen sensors based on BaCe_{0.7}Zr_{0.1}Y_{0.2}O_{3-δ} proton-conducting ceramic. *Sens. Actuators B Chem.* **2016**, *231*, 175–182. [[CrossRef](#)]
7. Li, Y.; Li, X.; Tang, Z.; Wang, J.; Yu, J.; Tang, Z. Potentiometric hydrogen sensors based on yttria-stabilized zirconia electrolyte (YSZ) and CdWO₄ interface. *Sens. Actuators B Chem.* **2016**, *223*, 365–371. [[CrossRef](#)]
8. Lee, E.B.; Hwang, I.-S.; Cha, J.-H.; Lee, H.-J.; Lee, W.-B.; Pak, J.J.; Lee, J.-H.; Ju, B.-K. Micromachined catalytic combustible hydrogen gas sensor. *Sens. Actuators B Chem.* **2011**, *153*, 392–397. [[CrossRef](#)]
9. Nishibori, M.; Shin, W.; Izu, N.; Itoh, T.; Matsubara, I.; Yasuda, S.; Ohtani, S. Robust hydrogen detection system with a thermoelectric hydrogen sensor for hydrogen station application. *Int. J. Hydrog. Energy* **2009**, *34*, 2834–2841. [[CrossRef](#)]
10. Baek, D.H.; Kim, J. MoS₂ gas sensor functionalized by Pd for the detection of hydrogen. *Sens. Actuators B Chem.* **2017**, *250*, 686–691. [[CrossRef](#)]
11. Lu, C.T.; Lin, K.W.; Chen, H.I.; Chuang, H.M.; Chen, C.Y.; Liu, W.C. A new Pd-oxide-Al_{0.3}Ga_{0.7}As MOS hydrogen sensor. *IEEE Electron. Device Lett.* **2003**, *24*, 390–392. [[CrossRef](#)]
12. Han, C.H.; Han, S.D.; Singh, I.; Toupance, T. Micro-bead of nano-crystalline F-doped SnO₂ as a sensitive hydrogen gas sensor. *Sens. Actuators B Chem.* **2005**, *109*, 264–269. [[CrossRef](#)]
13. Ippolito, S.J.; Kandasamy, S.; Kalantar-Zadeh, K.; Wlodarski, W. Hydrogen sensing characteristics of WO₃ thin film conductometric sensors activated by Pt and Au catalysts. *Sens. Actuators B Chem.* **2005**, *108*, 154–158. [[CrossRef](#)]

14. Dey, A. Semiconductor metal oxide gas sensors: A review. *Mater. Sci. Eng. B Solid-State Mater. Adv. Technol.* **2018**, *229*, 206–217. [CrossRef]
15. Paknahad, M.; Bachhal, J.S.; Ahmadi, A.; Hoorfar, M. Characterization of channel coating and dimensions of microfluidic-based gas detectors. *Sens. Actuators B Chem.* **2017**, *241*, 55–64. [CrossRef]
16. Yan, J.; Guo, X.; Duan, S.; Jia, P.; Wang, L.; Peng, C.; Zhang, S. Electronic nose feature extraction methods: A review. *Sensors* **2015**, *15*, 27804–27831.
17. Hastie, T.; Tibshirani, R.; Friedman, J.; Franklin, J. The elements of statistical learning: Data mining, inference and prediction. *Math. Intell.* **2005**, *27*, 83–85.
18. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
19. Cao, L.J.; Chua, K.S.; Chong, W.K.; Lee, H.P.; Gu, Q.M. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing* **2003**, *55*, 321–336. [CrossRef]
20. Yin, Y.; Tian, X. Classification of Chinese drinks by a gas sensors array and combination of the PCA with Wilks distribution. *Sens. Actuators B Chem.* **2007**, *124*, 393–397.
21. Bengio, Y. *Learning Deep Architectures for AI*; Now Publishers Inc.: Boston, MA, USA, 2009.
22. Baldi, P. Autoencoders, unsupervised learning, and deep architectures. *Proc. ICML Workshop Unsupervised Transf. Learn.* **2012**, *27*, 37–49.
23. Zhao, W.; Meng, Q.-H.; Zeng, M.; Qi, P.-F. Stacked sparse auto-encoders (SSAE) based electronic nose for Chinese liquors classification. *Sensors* **2017**, *17*, 2855.
24. Ma, D.; Gao, J.; Zhang, Z.; Zhao, H. Gas recognition method based on the deep learning model of sensor array response map. *Sens. Actuators B Chem.* **2021**, *330*, 129349.
25. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
26. Yan, K.; Zhang, D. Calibration transfer and drift compensation of e-noses via coupled task learning. *Sens. Actuators B Chem.* **2016**, *225*, 288–297.
27. Yi, R.; Yan, J.; Shi, D.; Tian, Y.; Chen, F.; Wang, Z.; Duan, S. Improving the performance of drifted/shifted electronic nose systems by cross-domain transfer using common transfer samples. *Sens. Actuators B Chem.* **2021**, *329*, 129162.
28. Barriault, M.; Alexander, I.; Tasnim, N.; O'Brien, A.; Najjaran, H.; Hoorfar, M. Classification and Regression of Binary Hydrocarbon Mixtures using Single Metal Oxide Semiconductor Sensor With Application to Natural Gas Detection. *Sens. Actuators B Chem.* **2021**, *326*, 129012. [CrossRef]
29. Gamboa, J.C.R.; da Silva, A.J.; Araujo, I.C.S. Validation of the rapid detection approach for enhancing the electronic nose systems performance, using different deep learning models and support vector machines. *Sens. Actuators B Chem.* **2021**, *327*, 128921.
30. Ramezankhani, M.; Crawford, B.; Narayan, A.; Voggenreiter, H.; Seethaler, R.; Milani, A.S. Making costly manufacturing smart with transfer learning under limited data: A case study on composites autoclave processing. *J. Manuf. Syst.* **2021**, *59*, 345–354. [CrossRef]
31. Finn, C.; Abbeel, P.; Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1126–1135. Available online: <http://arxiv.org/abs/1703.03400> (accessed on 18 July 2017).
32. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* **2014**, *4*, 3320–3328.
33. Neyshabur, B.; Sedghi, H.; Zhang, C. What is being transferred in transfer learning? *arXiv* **2020**, arXiv:2008.11687.
34. Ramezankhani, M.; Narayan, A.; Seethaler, R.; Milani, A.S. An Active Transfer Learning (ATL) Framework for Smart Manufacturing with Limited Data: Case Study on Material Transfer in Composites Processing. In Proceedings of the 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Victoria, BC, Canada, 10–12 May 2021; pp. 277–282. [CrossRef]
35. Hall, J.E.; Hooker, P.; Jeffrey, K.E. Gas detection of hydrogen/natural gas blends in the gas industry. *Int. J. Hydrog. Energy* **2020**, *46*, 12555–12565. [CrossRef]
36. Montazeri, M.M.; De Vries, N.; Afantchao, A.; Mehrabi, P.; Kim, E.; O'Brien, A.; Najjaran, H.; Hoorfar, M.; Kadota, P. A sensor for nuisance sewer gas monitoring. In Proceedings of the IEEE Sensors, Glasgow, UK, 29 October–1 November 2017; Volume 2017, pp. 1–3. [CrossRef]
37. Ahmadou, D.; Laref, R.; Losson, E.; Siadat, M. Reduction of drift impact in gas sensor response to improve quantitative odor analysis. In Proceedings of the IEEE International Conference on Industrial Technology (ICIT), Toronto, ON, Canada, 22–25 March 2017; pp. 928–933. [CrossRef]
38. Paknahad, M.; Ahmadi, A.; Rousseau, J.; Nejad, H.R.; Hoorfar, M. On-Chip Electronic Nose For Wine Tasting: A Digital Microfluidic Approach. *IEEE Sens. J.* **2017**, *17*, 4322–4329. [CrossRef]
39. Samarasekara, P. Hydrogen and methane gas sensors synthesis of multi-walled carbon nanotubes. *Chin. J. Phys.* **2009**, *47*, 361–369.
40. Park, N.-H.; Akamatsu, T.; Itoh, T.; Izu, N.; Shin, W. Calorimetric thermoelectric gas sensor for the detection of hydrogen, methane and mixed gases. *Sensors* **2014**, *14*, 8350–8362.
41. Brahim-Belhouari, S.; Bermak, A.; Shi, M.; Chan, P.C.H. Fast and robust gas identification system using an integrated gas sensor technology and Gaussian mixture models. *IEEE Sens. J.* **2005**, *5*, 1433–1444.

-
42. Westerwaal, R.; Gersen, S.; Ngene, P.; Darneveil, H.; Schreuders, H.; Middelkoop, J.; Dam, B. Fiber optic hydrogen sensor for a continuously monitoring of the partial hydrogen pressure in the natural gas grid. *Sens. Actuators B Chem.* **2014**, *199*, 127–132.
 43. Blokland, H.; Sweelssen, J.; Isaac, T.; Boersma, A. Detecting hydrogen concentrations during admixing hydrogen in natural gas grids. *Int. J. Hydrog. Energy* **2021**, *46*, 32318–32330.
 44. Ashkarran, A.A.; Olfatbakhsh, T.; Ramezankhani, M.; Crist, R.C.; Berrettini, W.H.; Milani, A.S.; Pakpour, S.; Mahmoudi, M. Evolving magnetically levitated plasma proteins detects opioid use disorder as a model disease. *Adv. Healthc. Mater.* **2020**, *9*, 1901608.
 45. Li, Q.; Gu, Y.; Wang, N.-F. Application of random forest classifier by means of a QCM-based e-nose in the identification of Chinese liquor flavors. *IEEE Sens. J.* **2017**, *17*, 1788–1794.
 46. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
 47. Ramezankhani, M.; Crawford, B.; Khayyam, H.; Naebe, M.; Seethaler, R.; Milani, A.S. A multi-objective Gaussian process approach for optimization and prediction of carbonization process in carbon fiber production under uncertainty. *Adv. Compos. Hybrid Mater.* **2019**, *2*, 444–455. [[CrossRef](#)]
 48. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.