RESEARCH ARTICLE

Deep Learning Framework With Essential Pre-Processing Techniques for Improving Mixed-Gas Concentration Prediction

MOONJUNG EO[1,*], JEONGYUN HAN[2,*], AND WONJONG RHEE[1,3,4], (Fellow, IEEE)

[1]Department of Intelligence and Information, Seoul National University, Gwanak-gu, Seoul 08826, South Korea
[2]Future Education Research Division, Center for Open Middle and High Schools, Jincheon-gun, Chungcheongbuk-do 27873, South Korea
[3]Interdisciplinary Program in Artificial Intelligence, Seoul National University, Gwanak-gu, Seoul 08826, South Korea
[4]AI Institute, Seoul National University, Gwanak-gu, Seoul 08826, South Korea

Corresponding author: Wonjong Rhee (wrhee@snu.ac.kr)

*Moonjung Eo and Jeongyun Han contributed equally to this work.

ABSTRACT Multiple gas detection in mixed-gas environments is a challenging issue in many engineering industries because some of the gases can raise defect rates and reduce production efficiency. For chemoresistive gas sensors, a precise estimation can be challenging because of the measurement variance and non-linear nature of the gas sensors, especially in a low concentration environment. A simple application of the deep learning models, however, does not yield sufficiently accurate predictions of the concentrations of multiple gases in gas mixtures; thus, it is essential to develop basic strategies for enhancing the accuracy in all possible ways. In this study, we develop a deep learning framework for achieving high accuracy of gas concentration prediction by studying the essential pre-processing techniques, learning task design, and architecture design. For the pre-processing, we study several aspects of processing time-series sensor data and identify the key techniques for complementing deep learning models' limitations. We utilize the mixed-gas nature for the learning task design and show that multi-task learning can generate a synergistic effect. Additionally, we show that a further improvement is possible by considering on-off classification as a part of the hybrid learning task. Concerning architecture design, we investigate Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) models after applying the identified pre-processing techniques. CNN outperformed other models in a joint analysis with the learning task. The effectiveness of our framework is confirmed with the UCI gas mixture dataset acquired using a chemical detection platform where 16 chemical sensors are exposed to ethylene, CO, and methane gases. Using the dataset, we study the basic techniques that can be effective to mixed-gas prediction. For the UCI dataset, our deep learning framework achieves a significant improvement in estimation accuracy when compared to the previous studies.

INDEX TERMS Chemical sensors, gas concentration prediction, deep neural network, pre-processing techniques, mixed-gas framework, hybrid-task.

I. INTRODUCTION

Detection of multiple gases in mixed-gas environments is an essential problem in many engineering fields because certain

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

gases can increase the defect rate and decrease production efficiency. Various methods have been proposed for the detection based on the chemical and physical properties of gases. For instance, metal oxide semiconductor sensors are promising components of modern gas detection devices owing to their advantages in sensitivity, cost, size, and speed

of analysis [1], [2]. However, these sensors have limitations, such as the lack of stability over time (also known as sensor drift) [3], [4], [5], [6] and cross-sensitivity to multiple gases [2], [7]. Because of such limitations, modeling of gas concentration prediction using these sensors is quite difficult, and advanced and customized modeling strategies are highly desired. In addition, the gases are usually released as mixtures of multiple gases in practical settings; thus, a more complex and challenging problem than building a simple model targeting a single gas must be solved. Many efforts have been made to address these limitations by developing sophisticated gas detection algorithms. In the early stages, traditional machine learning algorithms were used for gas detection, and more complex models were proposed to achieve better performance [2], [8], [9]. Recently, deep neural models have been actively utilized, and studies are being conducted to improve their performance through the development of novel and effective model structures [10], [11], [12].

In building these high-performing models, pre-processing plays an essential role, as well as novel model structures. Without proper pre-processing procedures, a model is usually not adequately trained to perform the target task. In the case of conventional benchmark datasets used in vision and language fields, typical pre-processing such as simple normalization can yield sufficient performance. However, it is unclear which pre-processing techniques should be explored and used for other application datasets obtained in an actual industrial environment. In the field of mixed-gas prediction, the previous studies [4], [13] are limited in that they have not carefully examined the pre-processing techniques that are specific to the nature of the mixed-gas data. Therefore, we focus on carefully studying a dataset for the purpose of developing specific pre-processing techniques for the mixed-gas prediction field. As we will show later, the identified pre-processing techniques turn out to be essential for improving prediction accuracy.

In deep learning, the success of representation learning or feature learning is heavily dependent on the choice of the learning task. For instance, a cross-entropy loss is typically chosen for classification tasks, and a mean square loss is chosen for regression tasks [14]. At the same time, it is well known that a learning task does not need to be a single goal. Choosing multiple goals is called multi-task learning. According to [15], multi-task learning is an approach to inductive transfer that improves generalization by using the domain information contained in the training signals of related tasks as an inductive bias. According to the findings of various studies [16], [17], [18], learning multiple related tasks simultaneously is superior to learning each task separately because the information learned for each task can help improve the learning of the other tasks. Thus, in the present study, we considered the concentrations of multiple gases as a multi-task. In addition, we formulated on-off detection as a classification task so that it could be utilized as a part of multiple tasks. This is called a hybrid task in our work. Although the multi-task learning framework

has been successfully adopted in various fields, only a few studies on gas detection and concentration prediction have been reported. In particular, to the best of our knowledge, no studies reported training of a multi-objective model that predicts the concentrations of multiple gases in gas mixtures using a single model. Another important factor for performance improvement in deep learning is architecture design. While the research community is still investigating how the deep neural network (DNN) works, it is well known to the practitioners that performance can be significantly affected by choice of architecture. For time series datasets, the common wisdom is to use RNN-based models; therefore, RNN and its variants have been applied most preferentially to gas identification for concentration prediction tasks [19], [20], [21]. At the same time, recent research has shown that CNN models learn more stably and that their performance is higher for datasets with sequential features, especially when a particular pattern exists in the data [10], [11]. We compared the performance of MLP, CNN, and RNN models to determine the most suitable architecture for our dataset. In addition, to verify the effect of model size and depth for our mixed-gas concentration dataset, we compared the performance of two models, ResNet-8 and ResNet-18. Based on the results of the studies mentioned above, we built deep-learning models using a CNN architecture. We showed they could significantly outperform the previously reported models for the UCI gas mixture dataset.

The contributions of this study are as follows:

- The effects of various pre-processing techniques (resampling, regularization, moving average, and delay correction) on the deep learning algorithm was investigated in detail
- Performance-enhancing learning tasks were designed to predict the concentrations of multiple gases in mixed gases environments or conduct hybrid tasks (regression and classification) in a single model, which further improved the algorithm's performance.
- The performance of three representative deep learning models (MLP, CNN, RNN) for time-series gas sensor datasets were compared, and the CNN model showed the best performance.
- Based on a suitable pre-processing procedure, our CNN-based algorithms with newly designed learning tasks outperform the previous model presented in [19] by 75.31-90.63%.

II. BACKGROUNDS

The present study focused on two aspects: pre-processing techniques for deep learning and deep-learning-based modeling. Data cleansing, data reduction, normalization, noise reduction, and lag time correction were chosen as major pre-processing techniques for deep learning of mixed-gas prediction. In this section, we summarize each pre-processing technique in detail. Deep-learning-based modeling can be divided into learning task design and architecture design.

Many studies on mixed gas detection or gas concentration prediction based on deep learning mainly focused on developing architecture design, and few works focused on learning task design. Therefore, we introduce the studies on learning task design and then summarize research results related to architecture design.

A. PRE-PROCESSING

1) DATA CLEANSING

Data cleansing consists of filling in missing values, identifying or removing outliers, and resolving inconsistencies [22]. Often raw data can confuse training procedures, resulting in unreliable modeling output. In the case of chemical gas sensor datasets, broken sensor data and settling time regions must be cleaned. Broken sensor data is recorded owing to sensor defects or errors in the data acquisition process. The settling time is the amount of time required for the chemical sensor to stabilize its output when it is turned on (and heating). In other words, it is a time interval during which the sensor is not adequately executing its function. Visualization methods (scatter plot, box plot, etc.) and statistical methods (standard deviation analysis, interquartile range analysis, etc.) can be used for the detection of broken sensor data and settling time regions [23], [24]. Certain algorithms, such as Potter's wheel or Intelliclean [25], can be used to identify irregular data points and precisely correct them. In the present study, we verified the settling time region and broken sensor data by visualizing the degree of change in the sensor values and removing these points before modeling.

2) DATA REDUCTION

As defined in [26], *data reduction is a process that obtains a reduced representation of the dataset that is much smaller in volume but yet produces the same (or almost the same) analytical results*. Thus, data reduction is essential for the efficient training of DNNs. Basic methods of data reduction include clustering, data aggregation, and downsampling. Time-series clustering is used to organize data points into groups based on their similarity. The objective is to maximize data similarity within clusters and minimize it across clusters [27]. Data aggregation is a type of multidimensional aggregation in which the original dataset is represented by aggregation at various levels of a data cube [28]. Downsampling is resampling a time-series dataset to a wider time frame [28], which was used in this study.

3) NORMALIZATION

According to [29], data normalization is a method that adjusts values measured on different scales to a notionally common scale. In references [30] and [31], owing to the wide range of resistance values collected by each sensor, if the chemical sensor data are properly normalized, it can help to make more reliable predictions of future test data, even when they are outside the range of the training dataset. There are many methods for data normalization,

such as z-score normalization, min-max normalization, and baseline normalization. Although z-score normalization is widely used [32], this method cannot effectively handle non-stationary time-series data. Therefore, we considered the min-max and baseline normalizations in the present study. The formula of the min-max normalization is shown in Equation 1, where X is the sensor data of the time series, max(X) is the maximum value of a sequence, and min(X) is the minimum value of a sequence. For every feature, the minimum value is transformed into 0, the maximum value is transformed into 1, and every other value is transformed into a decimal between 0 and 1. The baseline method is shown in Equation 2, where $X_{Standard}$ is the value obtained after subtracting the baseline treatment from $X_{Response}$, which is the true value of the sensor during the response, and $X_{Baseline}$ is the baseline value of the sensor in the air or standard gas. The ''standard'' data can effectively eliminate the impact on the environment and minimize environmental errors [2].

$$X_{normalized} = \frac{X - min(X)}{max(X) - min(X)} \quad (1)$$

$$X_{standard} = X_{Response} - X_{Baseline} \quad (2)$$

4) NOISE REDUCTION

Undesirable noise and disturbances in the remotely sensed time-series data must be minimized to preserve the original trends. Among noise reduction methods, moving average and binning-based clustering are particularly popular [26]. Binning methods smooth the sorted data value by consulting the neighborhood of each value, that is, the values around it [26]. The moving average is a method to analyze data points by creating a series of averages of different subsets of the full data set [23], [33]. Because the moving average is a simple and effective technique, it is most commonly used with time-series data to smooth out the short- and long-term effects [34]. We also used the moving average for noise reduction.

5) LAG TIME CORRECTION

The data from the chemical gas sensors have ''lag time'', which is the delay between the moment of gas injection and the acquisition of the sensor value. Both the response time of the sensor and the acquisition time are affected by many factors, including chemisorption and physisorption, cumulative exposure to the target gas and interfering gases, and maintenance [35]. The lag time correction is a time shift correction method that calculates the ''true'' profile from the measured resistance and actual gas concentration profiles based on the sensor time constant [36]. Therefore, it is necessary to compensate for this delay; moreover, drift correction has also been reported to improve the performance of the model [37], [38]. In this study, we set the lag time as a hyper-parameter and determined the proper value by analyzing the validation accuracy changes.
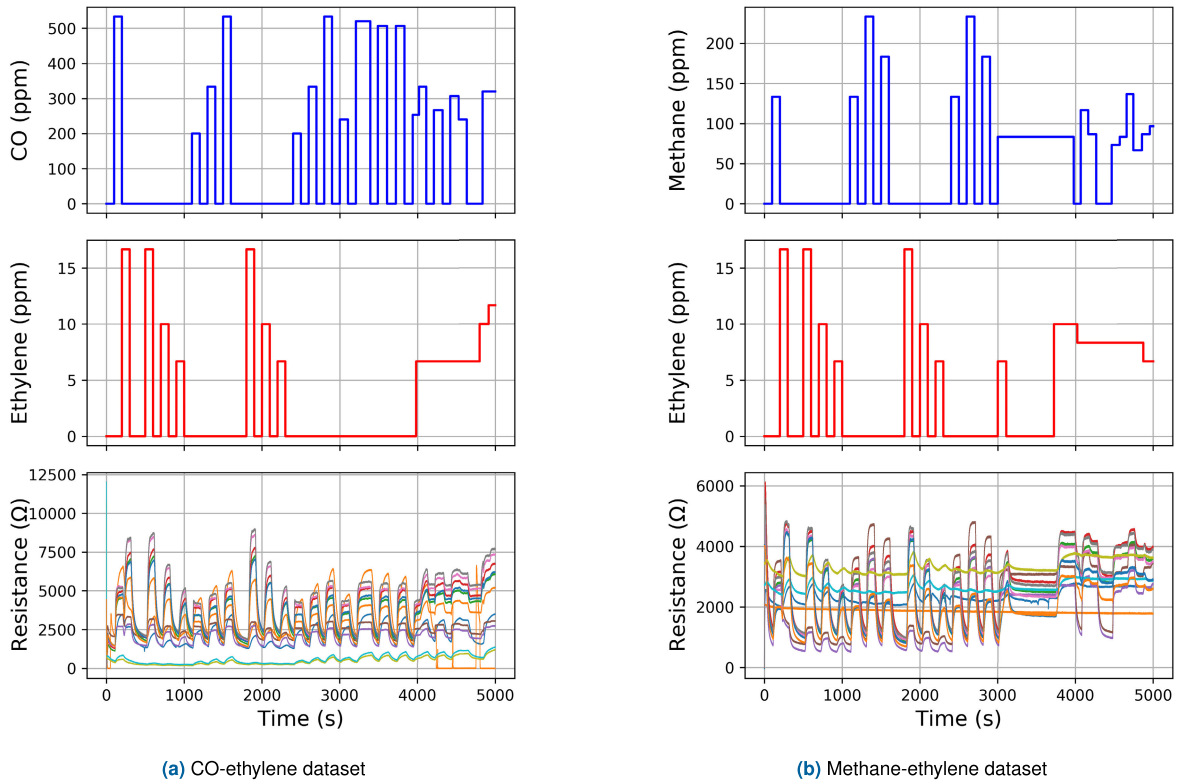
FIGURE 1. Example cases of the ground-truth concentration values of (a) CO and ethylene (b) methane and ethylene of the UCI dataset are shown together with the corresponding resistance measurements. The measurements are obtained using 16 sensors.

B. DEEP-LEARNING-BASED MODELING FOR GAS DETECTION AND CONCENTRATION PREDICTION

1) LEARNING TASK DESIGN

When training a deep-learning-based algorithm, it is necessary to define and design a task that is aligned with the ultimate purpose of learning. Many researchers have defined the task of predicting the gas concentration as a basic regression, and the algorithms were trained to minimize the corresponding loss [4], [10], [11], [39], [40], [41], [42]. Few studies have been conducted on the design of additional tasks. However, according to the findings of studies in various other fields, even if the target task is regression, the algorithm performance is sometimes improved when multi-task learning is performed by including an additional task, such as classification [43], [44], [45], [46]. This occurs because intrinsically valuable information can be obtained when performing classification and regression tasks, resulting in a synergistic effect on the algorithm's performance. To the best of our knowledge, the reference [13] is the only study that applied multi-tasking to the gas concentration prediction problem. Therefore, we used various learning tasks for the algorithm and attempted to determine their impact.

2) ARCHITECTURE DESIGN

Early research using deep learning was based on RNN-like models such as reservoir computing [19] and gated recurrent

unit neural networks [20]. More complex RNN models with better performance have been proposed in [14], [28], [45], and [47]. Although RNN and its variants architectures are regarded as suitable structures for sequence modeling [40], CNN is also an attractive solution for time-series datasets, and it has recently shown better performance than RNN architectures [47], [48], [49], [50]. An empirical study showed that CNNs could outperform RNNs across a diverse range of time-series modeling tasks and datasets [51]. In references [10] and [11], they proposed simple CNN-based models to classify gases in a mixed-gas environment and demonstrated faster learning than RNN-based models with comparable classification performance. A novel one-dimensional CNN was also proposed to predict the concentration of each gas in mixed-gas environments [52]. Although both classification and quantification of multiple gases in mixed-gas environments are required in an actual field, only a few studies have investigated these processes together.

III. EXPERIMENTS

A. DATASET DESCRIPTION

The experimental dataset used in this study is the "Gas sensor array under dynamic gas mixtures Data Set" from the public UCI gas mixture dataset. This UCI gas mixture dataset was collected using a chemical detection platform, where 16 chemical sensors were exposed to gas mixtures

at varying concentrations. The dataset was donated to the UCI machine learning repository (https://archive.ics.uci.edu) in 2014 and has since been used in several studies (e.g., [53], [54], [55]), including the benchmark for this study [19]. The UCI public dataset was produced for the study in [19], and it contains two sets of records for two different gas mixtures: CO-ethylene and methane-ethylene. Each sub-dataset consists of the records of 16 chemo-resistive gas sensors for the corresponding gas mixture and the measurement was continued for 12 hours. Visualization examples of the resistance measurements with the corresponding ground-truth concentration values are shown in FIGURE 1. It can be observed that similar patterns with randomness in release time and gas concentration are repeated over the collection period. The gas concentration level was randomly adjusted within the range of 0-20 ppm for ethylene, 0-600 ppm for CO, and 0-300 ppm for methane. Each time, the gas concentration level was kept constant in for a randomly chosen duration and transition time. The sensor array included 16 chemical sensors (Figaro Inc., US[1]) of 4 different types: TGS-2600, TGS-2602, TGS-2610, and TGS-2620 (4 units of each type). The sensors were integrated with customized signal conditioning and control electronics. The operating voltage of the sensors, which controls the operating temperature of the sensors, was kept constant at 5V throughout the entire experiment. The responses of the chemo-resistive gas sensors to the gases, namely, the changes in electric resistance, were collected at 100Hz resolution. It should be noted that the gas concentration was measured at the exit nozzle where the gas was released and not at the tunnel where the gas sensors actually met the gas. Further details are available in the original data description.[2]

B. METHODS
1) DATA PRE-PROCESSING
Multiple common pre-processing techniques were sequentially applied to obtain a clean dataset for building a deep-learning-based gas detection and concentration prediction model.

Data cleansing To clean the sensor dataset, a visualization was first used to understand the overall sensor responses and identify abnormal patterns that needed to be cleaned. As shown in FIGURE 2a, the sensors appear to have an unstable initialization phase when they are turned on. Therefore, the data points collected during this phase were removed. We investigated every sensor's initial responses to determine the initialization time required to cut the unstable data points from the experimental dataset. For example, the shaded area in FIGURE 2a was deemed unstable because the sensors' responses appear to be converging to a stable phase during the initialization. After investigating

[1]5400 Newport Drive, Suite 19, Rolling Meadows, IL 60008 USA, Figaro USA, Inc.
[2]Dataset details can be found at the following link: https://archive.ics.uci.edu/ml/datasets/gas+sensor+array+under+dynamic+gas+mixtures
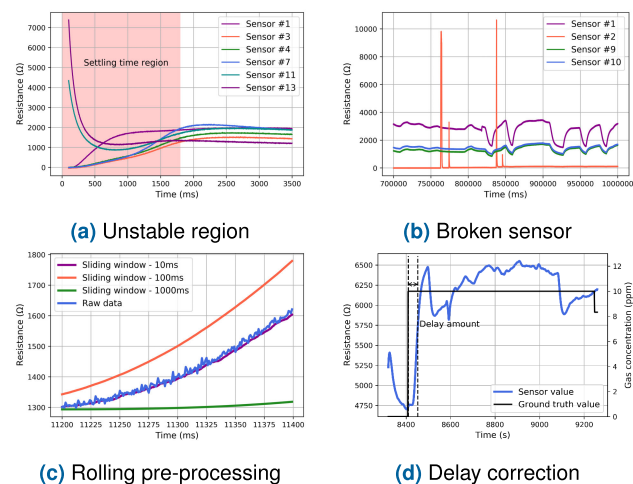

FIGURE 2. Visualization of the CO-ethylene dataset. The needs for specific pre-processing techniques can be identified. (a) Sensor response in the first few seconds of measurement. The sensors show unstable and abnormal responses in the shaded area. (b) Sensor response for varying gas concentration. Sensor #2 shows an abnormal pattern when compared to the other sensors. (c) The effect of window size (CO gas; sensor #1). As the sliding window size is increased, the noise signal is reduced. However, the magnitude is severely distorted when the window size is excessively large. (d) The ground-truth of the injected CO concentration and the response of sensor #1. There is a delay between the two measurements, indicating that a lag time correction can be helpful.

the initial response of each sensor, initial 2000 data points (approximately 2000 ms) were removed from all sensors. In addition, as shown in FIGURE 2b, a faulty sensor (sensor 2) can be easily distinguished from the other sensors using a simple visualization method because sensor 2 shows a significantly different pattern in resistance values compared to the other sensors. Thus, the data from sensor 2 was excluded from this study.

Data reduction It is infeasible to analyze a sufficient number of deep-learning-based models with a high-resolution (100 Hz) sensor dataset because of the associated high computational cost. For this reason, data reduction is essential when using deep learning; however, it is also challenging to find an appropriate level of reduction that downsizes the dataset sufficiently but does not eliminate too much information from the original dataset. We adopted a two-step approach to solve this problem: visualization and grid search. First, by visualizing each sensor's measurements at various downsampling rates, we determined a suitable search interval for the grid search (0.1, 1, 10, and 100 s), which was used in the next step. Second, a grid search was performed at the chosen interval by considering the effect of each parameter on model performance. Ultimately, 1 s was selected as the optimal downsampling rate for our dataset because it did not seriously harm the model performance.

Normalization This study considered two types of normalization methods: min-max and baseline. While min-max normalization has no hyperparameter to optimize, baseline normalization requires a window size to set the data range to calculate the $X_{Baseline}$ used in Equation 2. To determine
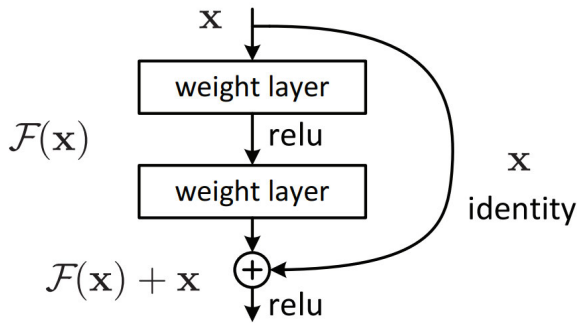
FIGURE 3. The basic residual block of ResNet structure from the original ResNet study (Figures were adapted from FIGURE 2 of the reference [56]). The structure contains skip (shortcut) connections to enable the gradient to flow directly to the bottom layers.

an appropriate window size, a grid search method was used with intervals of 200, 400, 600, 800, and 1000 s, yielding 1000 s for the CO-ethylene dataset and 600 s for the methane-ethylene dataset based on the model validation performance.

Noise reduction In this study, a moving average method was used for noise reduction. The appropriate window size for this method was determined using a grid search method. As shown in FIGURE 2c, with increasing window size, the smoothness of the output increases, and the sharp modulations become increasingly blunt, but at the same time, the data information can be distorted if the chosen window size is too large. The modeling performances were compared for the window sizes of 0, 10, 100, and 1000 ms. Window sizes of 100 and 10 were selected for CO-ethylene and methane-ethylene data, respectively.

Lag time correction In the experimental dataset, the gas concentration was measured at the exit nozzle rather than at the tunnel, where the sensors actually met the gas. This means that the gas needs a certain amount of time to spread sufficiently in the tunnel so that the sensor responds to the gas. To properly match the sensor responses and the actual ground truth of gas concentration, one of the signals needs to be shifted by the lag time, which is called lag time correction. For this, we first drew plots (FIGURE 2d) to determine a reasonable range for lag time. Considering the validation performance of our models, 47 s was selected as the optimal lag time.

2) DESIGN OF LEARNING TASK

The CNN-based ResNet-8 was used as the base backbone model, and based on it, four types of variations depending on the learning task design were introduced: single-gas basic model, mixed-gas basic model, single-gas hybrid model, and mixed-gas hybrid model. The ResNet model is comprised of multiple CNN modules and extends neural networks to a deep structure by adding a shortcut connection in each residual block to enable the gradient to flow directly through the network to the bottom layers (FIGURE 3). As shown

in FIGURE 4a, the single-gas model was built by adding dense (fully connected) and regression output layers to the backbone model (ResNet-8). We refer to this single-gas model as the basic model in the present study. For the mixed-gas model, as shown in FIGURE 4b, three dense layers and a regression output layer were added at the end of the backbone model for each gas. The mean absolute error (MAE) loss function from Equation 3 was used for training regression of both the basic and multiple-gas models. We also considered a hybrid task model [13] that performs classification and a regression task in a single model to improve the performance of concentration estimation. Note that the classification task is introduced only to improve the regression task's performance. The regression task was used to forecast the gas concentration, and the classification task was used to determine whether gas was present. The ground truth of the classification task was set to be false when the gas concentration was zero and true otherwise. This corresponds to an on-off classification that can be a useful task for improving the learning of deep learning models. The regression task was performed using the MAE (Mean Absolute Error) loss function from Equation 3, and the classification task was performed using the cross-entropy loss function from Equation 4. A hybrid task can be used with both single- or multiple-gas models. As shown in FIGURE 4c, there are two outputs in the last layer to predict the gas concentration and presence in the single-gas model. For the multiple-gas hybrid models, as shown in FIGURE 4d, each branch (dense layer) has two outputs in the last layer for the hybrid tasks.

$$MAE = \frac{1}{N}\Sigma_{i=1}^{N}|Actual_i - Predicted_i| \qquad (3)$$

$$Cross\ Entropy = -\Sigma_{i=1}^{N}Actual_i \times log(Predicted_i) \qquad (4)$$

3) ARCHITECTURE DESIGN

Three types of DNNs are commonly used: MLP, CNN, and RNN. MLP is the most basic architecture of modern DNNs, and it usually serves as the baseline architecture of DNN algorithms. Unlike the fully connected layers in MLPs, convolution layers in CNN models extract simple features from the input by performing convolution operations that are computationally cheap and allow the adoption of many layers. CNN models can capture the high-level representation of input data, making them the most popular choice for computer vision tasks, such as image classification and object detection. RNNs have recurrent connections with hidden states. This looping constraint ensures that the sequential information is captured in the input data. Furthermore, each block in an RNN has an internal memory that stores the computational information from the previous time steps. As mentioned above, the characteristics used or extracted features when modeling data differ for each type of DNNs, affecting the performance of the model. A performance comparison determined the type of deep learning model best suited for modeling the time-series dataset with a specific
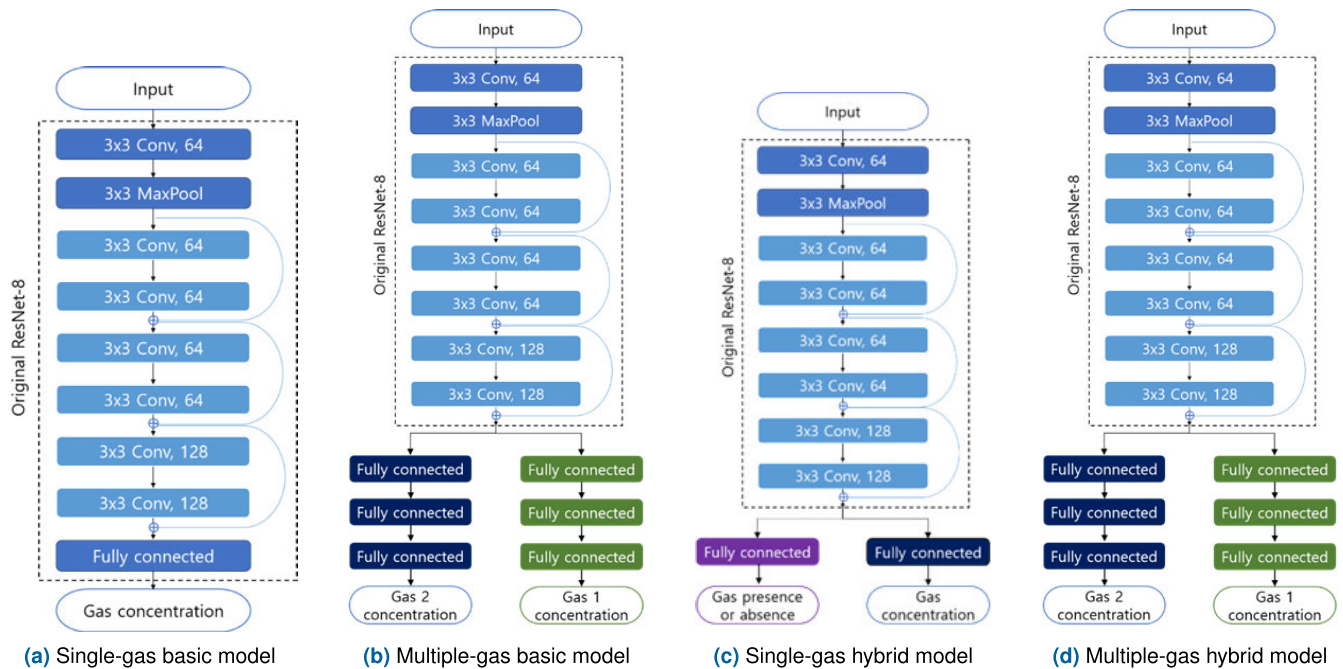
FIGURE 4. Deep-learning-based basic models and hybrid models for this study: (a) basic DNN to predict gas concentration, (b) the multiple-gas model to predict gas1 and gas2 concentrations simultaneously, (c) the hybrid model to predict both gas concentration and gas detection and (d) the mixed-gas hybrid model to detect gas1 and gas2 and predict their concentrations by using a model composed of 2 branches.

pattern utilized in this study. In addition, multiple reports have already demonstrated that, for the same model type, the model's performance improves with the model's depth [56], [57]. This tendency was verified by evaluating and comparing the performance of two ResNet models.

C. EXPERIMENTAL DETAILS
1) PRE-PROCESSING
In this study, the UCI public dataset was split into the train (70%), validation (10%), and test (20%). Among all the pre-processing techniques, data cleansing, and data reduction were applied by default, and the other pre-processing and modeling experiments were performed with them.

2) MODELLING
We used ResNet-8 as the backbone model, with several branch architectures stacked on top, to analyze both pre-processing techniques and learning task design. Each model's tasks were classified into four categories, which are described in FIGURE 4. For performance comparison depending on the model architecture design, we used 5-layer MLP (each layer was composed of 300 neurons), ResNet-8 (all layers were CNN modules), and LSTM (the most popular RNN model) with 330 neurons as the state vector. For a fair comparison, each model was designed to have a similar number of parameters. This is an essential requirement to prevent a model with a larger size from unfairly outperforming the others. ResNet-18, composed of 18 layers, was used as the deeper model with additional parameters.

3) LEARNING
A stochastic gradient descent optimizer was used for all tasks. Three values (0.1, 0.01, 0.001) were used for the learning rate, and the value yielding the highest model validation accuracy was selected (0.01). Early stopping was performed to avoid overfitting. The training was stopped when the validation performance did not change for more than 15 epochs.

4) PERFORMANCE MEASUREMENT
MAE (Mean Absolute Error) was utilized to evaluate the model performance because it is a standard performance metric that is resistant to outliers because it is a linear score that averages all individual differences with an equal weighting [58]. The average and standard deviation of three repeated experiments were used to evaluate the performance of the models.

IV. RESULTS
The model performances depending on the combination of data pre-processing methods and model structure are summarized in Table 1-4.

A. PRE-PROCESSING TECHNIQUES
The analysis results of the effect of each pre-processing on the performance of the model are as follows.

1) DOWNSAMPLING
For the time-series sensor data, a considerable amount of information is redundant, and many data points are simply

TABLE 1. Model performance on CO-ethylene data.

(a) Model performance on CO data.

Pre-processing techniques				Performance for each learning task			
Normalization		Moving	Lag time	Single-gas		Multiple-gas	
Min-max	Baseline	average	correction	Basic	Hybrid	Basic	Hybrid
Benchmark [19]				45.9			
Baseline (modeling with raw data)				24.60 ± 8.50	31.16 ± 4.91	10.27 ± 2.15	24.68 ± 6.40
V				22.76 ± 2.32	41.4 ± 21.51	10.62 ± 2.14	21.52 ± 0.26
	V			20.73 ± 0.94	19.36 ± 1.50	8.93 ± 1.32	17.69 ± 1.44
	V	V		15.96 ± 0.58	23.14 ± 5.76	9.43 ± 0.54	16.20 ± 2.38
	V		V	10.14 ± 1.95	15.42 ± 6.93	5.36 ± 1.86	**7.12 ± 0.43**
	V	V	V	**9.20 ± 0.44**	**9.48 ± 0.68**	**4.30 ± 0.49**	8.69 ± 1.18

(b) Model performances(MAE) on ethylene dataset.

Pre-processing techniques				Performance for each learning task			
Normalization		Moving	Lag time	Single-gas		Multiple-gas	
Min-max	Baseline	Average	correction	Basic	Hybrid	Basic	Hybrid
Benchmark [19]				1.7			
Baseline (modeling with raw data)				1.25 ± 0.20	1.95 ± 0.90	1.07 ± 0.34	1.39 ± 0.97
V				1.36 ± 0.24	1.45 ± 0.47	0.85 ± 0.11	0.94 ± 0.15
	V			1.02 ± 0.08	0.88 ± 0.04	0.93 ± 0.02	1.09 ± 0.34
	V	V		0.94 ± 0.22	0.78 ± 0.05	0.79 ± 0.04	0.95 ± 0.21
	V		V	0.46 ± 0.01	0.45 ± 0.18	**0.43 ± 0.01**	**0.40 ± 0.06**
	V	V	V	**0.43 ± 0.06**	**0.37 ± 0.08**	**0.43 ± 0.02**	**0.40 ± 0.34**

replicated versions of the previous points with noise. Down-sampling can reduce training time, making training more efficient. We reduced the amount of data by downsampling it to 1 Hz, which considerably sped up the optimization of various hyperparameters. We saved training time of 117 min (120 min with the original data and 2.4 minutes with the reduced data) per epoch.

2) NORMALIZATION
Normalization always reduces the standard deviation of the model performance, regardless of the method used. This is the case for both the CO-ethylene and ethylene-methane datasets (Tables 1 and 2, respectively). In addition, except for the single-gas target modeling on the ethylene-methane dataset, the normalization effectively improved the model performance. In most cases, baseline normalization was more effective than min-max normalization for improving the prediction performance on both datasets. In addition, in three repeated experiments, the standard deviation decreased in most cases after the normalization. Therefore, the normalization stabilized the training process.

3) MOVING AVERAGE
Moving average usually improved the model performance on our dataset (in 6 out of 8 cases on the CO-ethylene dataset and all cases on the ethylene-methane dataset, as shown in Tables 1 and 2, respectively). The moving average was helpful in predicting the concentration of ethylene in both datasets. The performance of single-gas

basic modeling for CO was improved by 23.01% in the CO-ethylene dataset, and the performance of single-gas basic modeling for ethylene increased by 31.80% in the ethylene-methane dataset. Even in the case when the performance was degraded, using both moving average and lag time correction resulted in a performance improvement.

4) LAG TIME CORRECTION
We conducted lag correction, which resulted in significant increases in performance (by approximately 50%) for all datasets and all predictive modeling setups. The effect of the lag correction is substantially larger in the ethylene-methane dataset than in the CO-ethylene dataset, and the performance was enhanced by up to 83.14% in the case of ethylene and 64.09% in the case of methane in the ethylene-methane dataset. For a more elaborate correction, the hyperparameters may be replaced with an algorithm such as a linear regression, which adapts the correction period as a predicted value. This method allows for a more sophisticated calibration of chemical sensors to drift over time, and it should be applied to long-term gas concentration prediction. In our study, we limited our interest to a constant correction due to the limited size of the UCI dataset. The application of the moving average and lag time correction together yields the best modeling performance in most cases (13 out of 16). As shown in Tables 1 and 2, the best performance in most cases is obtained when the moving average and lag correction are applied together.

TABLE 2. Model performances(MAE) on methane-ethylene dataset.

(a) Model performance on methane data

| Pre-processing techniques | | | | Performance for each learning task | | | |
| Normalization | | Moving | Lag time | Single-gas | | Multiple-gas | |
Min-max	Baseline	average	correction	Basic	Hybrid	Basic	Hybrid
Benchmark [19]				12.5			
Baseline (modeling with raw data)				11.00 ± 5.16	24.8 ± 23.33	6.72 ± 1.86	8.63 ± 1.93
V				19.03 ± 14.48	22.2 ± 6.63	11.85 ± 4.68	14.4 ± 14.32
	V			11.17 ± 2.94	8.71 ± 1.39	6.60 ± 1.58	6.20 ± 0.62
	V	V		10.06 ± 2.14	9.7 ± 2.40	5.99 ± 0.43	2.67 ± 1.23
	V		V	4.42 ± 0.51	**3.79 ± 0.08**	**2.37 ± 0.62**	2.63 ± 0.23
	V	V	V	**3.41 ± 0.17**	5.12 ± 1.58	2.53 ± 0.10	**2.53 ± 0.63**

(b) Model performance on ethylene data

| Pre-processing techniques | | | | Performance for each learning task | | | |
| Normalization | | Moving | Lag time | Single-gas | | Multiple-gas | |
Min-max	Baseline	average	correction	Basic	Hybrid	Basic	Hybrid
Benchmark [19]				1.3			
Baseline (modeling with raw data)				1.92 ± 0.48	2.57 ± 0.35	1.61 ± 0.79	2.00 ± 0.39
V				2.70 ± 0.24	2.83 ± 0.25	2.21 ± 0.11	2.59 ± 0.05
	V			2.61 ± 0.26	1.65 ± 0.43	1.11 ± 0.16	1.31 ± 0.38
	V	V		1.78 ± 0.17	1.44 ± 0.34	1.08 ± 0.08	1.12 ± 0.44
	V		V	0.44 ± 0.05	0.38 ± 0.16	0.47 ± 0.28	0.28 ± 0.12
	V	V	V	**0.42 ± 0.03**	**0.30 ± 0.02**	**0.32 ± 0.16**	**0.18 ± 0.06**

TABLE 3. Model performance (MAE) comparison.

(a) Model performance on CO-ethylene dataset

| Architecture | MAE for CO | | | | MAE for ethylene | | | |
| | Single-gas | | Multiple-gas | | Single-gas | | Multiple-gas | |
	Basic	Hybrid	Basic	Hybrid	Basic	Hybrid	Basic	Hybrid
MLP	10.70 ± 0.31	12.32 ± 1.81	9.77 ± 1.15	7.98 ± 0.57	0.66 ± 0.18	0.57 ± 0.05	0.69 ± 0.21	0.57 ± 0.13
ResNet-8	9.20 ± 0.44	9.48 ± 0.68	4.30 ± 0.49	8.69 ± 1.18	0.43 ± 0.06	0.37 ± 0.08	0.43 ± 0.02	0.40 ± 0.34
LSTM	16.39 ± 3.44	16.31 ± 1.04	12.87 ± 1.72	12.34 ± 0.57	0.40 ± 0.04	0.42 ± 0.10	0.50 ± 0.14	0.59 ± 0.14
ResNet-18	**8.98 ± 0.39**	**8.83 ± 0.41**	**4.21 ± 0.06**	**6.97 ± 0.57**	**0.29 ± 0.01**	**0.31 ± 0.05**	**0.36 ± 0.03**	**0.35 ± 0.01**

(b) Model performance on methane-ethylene dataset

| Architecture | MAE for methane | | | | MAE for ethylene | | | |
| | Single-gas | | Multiple-gas | | Single-gas | | Multiple-gas | |
	Basic	Hybrid	Basic	Hybrid	Basic	Hybrid	Basic	Hybrid
MLP	7.47 ± 1.51	9.58 ± 4.67	5.30 ± 1.36	5.40 ± 1.27	0.45 ± 0.04	0.54 ± 0.09	0.40 ± 0.08	0.63 ± 0.24
ResNet-8	3.41 ± 0.17	5.12 ± 1.58	2.53 ± 0.10	2.53 ± 0.63	0.42 ± 0.03	0.30 ± 0.02	0.32 ± 0.16	0.18 ± 0.06
LSTM	7.14 ± 0.23	5.64 ± 0.98	4.80 ± 0.95	4.48 ± 0.64	0.48 ± 0.92	0.44 ± 0.91	0.69 ± 0.23	0.39 ± 0.04
ResNet-18	**2.55 ± 0.39**	**2.94 ± 0.50**	**2.33 ± 0.68**	**2.39 ± 0.12**	**0.19 ± 0.04**	**0.20 ± 0.01**	**0.26 ± 0.04**	**0.18 ± 0.05**

B. LEARNING TASK DESIGN

According to the obtained results, multiple-gas target modeling performed better than single-gas modeling for all datasets except for ethylene in Table 1b (see bold numbers in Tables 1 and 2). For example, in the multiple-gas basic model, the maximum modeling performance for CO gas is 53.26% better than that in the single-gas basic model; for methane, the difference is 30.50%, and for ethylene in the ethylene-methane dataset, it is 23.81%. Furthermore, except for one case, the multiple-gas modeling outperforms the single-gas modeling throughout all pre-processing steps on both the CO-ethylene and ethylene-methane datasets. We also combined the hybrid task with ours, and the results are shown in the "hybrid" column in Tables 1 and 2. In the modeling for CO and methane, the regression-only task yielded better results, whereas the hybrid task yielded better results for ethylene in both tables (performance higher by 43.75% than for the basic task). The reason for hybrid learning's deteriorating effect on CO and methane and the improving effect on ethylene may be attributed to

TABLE 4. Comparison of benchmark and our best performance.

Target	CO-ethylene dataset		Methane-ethylene dataset	
	CO (MAE)	Ethylene (MAE)	Methane (MAE)	Ethylene (MAE)
Benchmark [19]	45.9	1.70	12.5	1.30
Ours	4.21	0.29	2.33	0.18
MAE reduction rate (%)	90.83%	82.94%	81.36%	86.15%
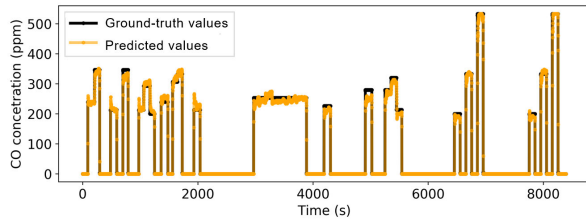
(a) CO gas results of CO-ethylene dataset
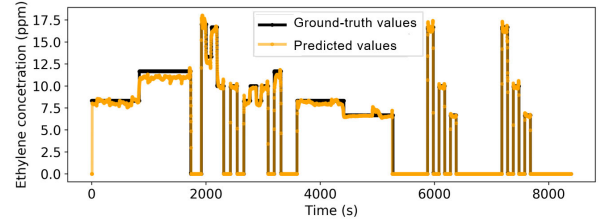
(b) Ethylene gas results of CO-ethylene dataset
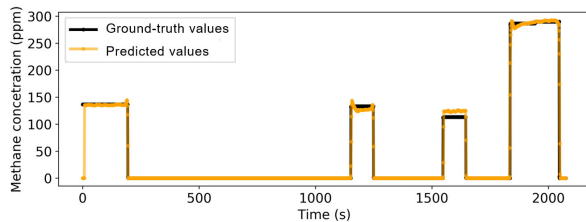
(c) Methane gas results of methane-ethylene dataset
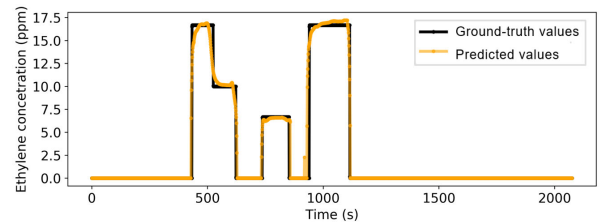
(d) Ethylene gas results of methane-ethylene dataset

FIGURE 5. Ground-truth values and predicted values for the test datasets.

the concentration range of each gas. Compared to CO or methane, ethylene was injected within a narrow concentration range. Therefore, the on-off classification was likely more useful for learning the regression output of ethylene than the output of CO or methane. For CO and methane with a wider concentration range, on-off classification did not help improve the representation learning of the regression task.

C. ARCHITECTURE DESIGN

To elucidate the benefits of using CNN-based architecture, we evaluated three different module-based models: ResNet-X for CNN-based architecture, MLP for multi-layer perceptron-based architecture, and LSTM for RNN-based architecture (Table 3). In addition, the performance of CNN-based models is known to improve with the number of layers; thus, we verified whether this property holds for the considered models (ResNet-8 vs. ResNet-18). As shown in Table 3, the ResNet models outperform MLP and LSTM models by a large margin. In addition, although LSTM was designed for time-series data, MLP models in certain cases (6 out of 16) outperform LSTM models. The ResNet-18 model, which contains more layers and parameters than the ResNet-8 model, consistently outperforms the ResNet-8 model in all cases. This result suggests that even larger models might be useful for improving performance further.

D. PERFORMANCE SUMMARIZATION AND COMPARISON
1) PERFORMANCE COMPARISON TO BENCHMARK
The main goal of this work is to investigate the essential factors for developing a deep learning framework for gas concentration prediction. While the MAE values themselves are not the most important results of our work, we provide a comparison with the benchmark provided in the UCI dataset's original paper [19]. As shown in Table 4, our model outperforms the benchmark for all datasets and gases. As shown in Table 1 (rows 1 and 2), the proposed deep learning models outperform the benchmark even without any pre-processing. With all the enhancements in our framework, the resulting method shows between 81.36% and 90.83% improved performance than the benchmark method.

2) VISUALIZATION OF THE PREDICTED VALUES
The concentration plots of the ground-truth and the predicted values are shown in Figure 5. For each plot, the best-performing model was used for generating the predicted values. It can be observed that the prediction accuracy is high in general but can deteriorate when the gas concentration changes rapidly. The result indicates that the chosen models are effective for predicting gas concentration, even though the training samples and the testing samples are from different time sequences.

V. DISCUSSION

A. DIFFERENCE IN THE LEVEL OF MAE PERFORMANCE

In our results, the level of MAE performance is strongly dependent on the type of gas. For instance, the MAE performance of our model in Table 4 is 4.21 for CO but only 0.29 or 0.18 for ethylene. The large difference can be explained by the scale of the injected gas. While CO was injected in the scale of 0 to 600ppm, ethylene was injected in a much smaller scale of 0 to 20ppm. When normalized by the maximum ppm values, the normalized MAE is 0.007 ($= 4.21/600$) for CO and 0.015 ($= 0.29/20$) or 0.009 ($= 0.18/20$) for ethylene. For methane, the normalized MAE is 0.008 ($= 2.33/300$). Therefore, it can be observed that the normalized MAE values for the dataset are comparable for the three gas types. This indicates that the chemo-resistive sensors and our model are capable of handling the lower ppm levels as well as the higher ppm levels, at least for the studied UCI public dataset.

B. GAS COMPOSITION IN REAL WORLD

We have focused on developing and understanding general strategies when predicting mixed-gas concentrations with deep learning models. The UCI dataset was constructed to cover all possible scenarios, including increasing or decreasing transitions and biased gas compositions between the two mixed gases, and therefore the dataset construction was adequate for our purpose of developing general strategies. In the real world, however, an adjustment according to the specific gas compositions in each situation might be necessary. For instance, a biased gas composition might be the dominant situation in a real world scenario. In such a scenario, each technique considered in our work might exhibit different behavior, and an adjustment might be necessary. In fact, it might be possible to achieve even larger improvements for such a biased scenario because the techniques can be tuned for the dominant situation.

C. POSSIBLE DIRECTIONS FOR A FURTHER IMPROVEMENT

In our study, we have shown that simple but adequate adjustments in pre-processing can have a significant influence when a deep learning model is used. While this insight should be useful and applicable for many mixed-gas prediction tasks, we were not able to specify the characteristic of the mixed-gas task that makes a particular pre-processing technique effective. This remains as a limitation of our work, and studying a variety of datasets collected with diverse sensor types and environments might reveal ways to achieve a further improvement. From the algorithm side, we have shown that learning task design and architecture design can have a substantial effect. While the overall gain is large, there can be other techniques that are impactful depending on the task characteristics. Considering that deep learning techniques are continuing to evolve, following the latest developments in advanced deep learning techniques would be a reasonable direction for a future improvement.

VI. CONCLUSION

Detecting and predicting the concentrations of various gases in mixed-gas environments is crucial in many fields of industry. Several attempts have been made to apply DNNs to gas concentration prediction problems. However, most studies did not carefully consider data pre-processing, which is a simple but crucial step in using deep learning algorithms. In addition, only a few studies predicted the concentrations of multiple gases in gas mixtures. We introduced and analyzed various pre-processing methods for chemical sensor data in this study. First, representative pre-processing methods, such as data cleansing, data reduction, noise reduction, normalization, and lag correction, were considered, and the effects of each on model performance were analyzed. Both normalization and delay correction was the most powerful methods for improving the model performance on our dataset. Second, we presented a novel mixed-gas framework for simultaneous multiple-gas detection and concentration estimation in a mixed-gas environment. To the best of our knowledge, the present study is the first to exploit the proposed framework for gas mixture classification and regression. In addition, we compared the performance of three different modules (MLP, RNN, and CNN) based on DNNs. The CNN module-based model performed exceptionally well, with the best results obtained using ResNet-18 with additional layers (an 81.36% improvement over a previous study). This study provides the most basic but essential insight into the use of various pre-processing techniques and multi-task-learning-based DNNs for mixed-gas detection and concentration estimation.

REFERENCES

[1] J. Kukkola, J. Mäklin, N. Halonen, T. Kyllönen, G. Tóth, M. Szabó, A. Shchukarev, J.-P. Mikkola, H. Jantunen, and K. Kordás, ''Gas sensors based on anodic tungsten oxide,'' *Sens. Actuators B, Chem.*, vol. 153, no. 2, pp. 293–300, Apr. 2011.

[2] L. Han, C. Yu, K. Xiao, and X. Zhao, ''A new method of mixed gas identification based on a convolutional neural network for time series classification,'' *Sensors*, vol. 19, no. 9, p. 1960, Apr. 2019.

[3] A. Rudnitskaya, ''Calibration update and drift correction for electronic noses and tongues,'' *Frontiers Chem.*, vol. 6, p. 433, Sep. 2018.

[4] A. Vergara, S. Vembu, T. Ayhan, M. A. Ryan, M. L. Homer, and R. Huerta, ''Chemical gas sensor drift compensation using classifier ensembles,'' *Sens. Actuators B, Chem.*, vol. 166, pp. 320–329, May 2012.

[5] L. Zhang and D. Zhang, ''Domain adaptation extreme learning machines for drift compensation in E-nose systems,'' *IEEE Trans. Instrum. Meas.*, vol. 64, no. 7, pp. 1790–1801, Jul. 2015.

[6] A. Ziyatdinov, S. Marco, A. Chaudry, K. Persaud, P. Caminal, and A. Perera, ''Drift compensation of gas sensor array data by common principal component analysis,'' *Sens. Actuators B, Chem.*, vol. 146, pp. 460–465, Sep. 2010.

[7] V. Krivetskiy, A. Efitorov, A. Arkhipenko, S. Vladimirova, M. Rumyantseva, S. Dolenko, and A. Gaskov, ''Selective detection of individual gases and CO/H$_2$ mixture at low concentrations in air by single semiconductor metal oxide sensors working in dynamic temperature mode,'' *Sens. Actuators B, Chem.*, vol. 254, pp. 502–513, Jan. 2018.

[8] B. Wang, W. Kong, and P. Zhao, "An air quality forecasting model based on improved ConvNet and RNN," *Soft Comput.*, vol. 25, no. 14, pp. 9209–9218, Jul. 2021.

[9] A. Borovykh, S. Bohte, and C. W. Oosterlee, "Conditional time series forecasting with convolutional neural networks," 2017, *arXiv:1703.04691*.

[10] G. Wei, G. Li, J. Zhao, and A. He, "Development of a LeNet-5 gas identification CNN structure for electronic noses," *Sensors*, vol. 19, no. 1, p. 217, 2019.

[11] P. Peng, X. Zhao, X. Pan, and W. Ye, "Gas classification using deep convolutional neural networks," *Sensors*, vol. 18, no. 2, p. 157, Jan. 2018.

[12] D. Ma, J. Gao, Z. Zhang, and H. Zhao, "Gas recognition method based on the deep learning model of sensor array response map," *Sens. Actuators B, Chem.*, vol. 330, Mar. 2021, Art. no. 129349.

[13] H. Liu, Q. Li, and Y. Gu, "A multi-task learning framework for gas detection and concentration estimation," *Neurocomputing*, vol. 416, pp. 28–37, Nov. 2020.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[15] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, pp. 41–75, Dec. 1997.

[16] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 121–135, Jan. 2019.

[17] K. Wu and G. W. Wei, "Quantitative toxicity prediction using topology based multitask deep neural networks," *J. Chem. Inf. Model.*, vol. 58, no. 2, pp. 520–531, 2018.

[18] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," 2016, *arXiv:1605.05101*.

[19] J. Fonollosa, S. Sheik, R. Huerta, and S. Marco, "Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring," *Sens. Actuators B, Chem.*, vol. 215, pp. 618–629, Aug. 2015.

[20] S. Wang, Y. Hu, J. Burgués, S. Marco, and S.-C. Liu, "Prediction of gas concentration using gated recurrent neural networks," in *Proc. 2nd IEEE Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, Aug. 2020, pp. 178–182.

[21] Q. Wang, H. Qi, and F. Liu, "Time series prediction of e-nose sensor drift based on deep recurrent neural network," in *Proc. Chin. Control Conf. (CCC)*, Jul. 2019, pp. 3479–3484.

[22] A. Gusrialdi, S. Hirche, D. Asikin, T. Hatanaka, and M. Fujita, "Voronoi-based coverage control with anisotropic sensors and experimental case study," *Intell. Service Robot.*, vol. 2, no. 4, pp. 195–204, Oct. 2009.

[23] L. Olsson and L. Eklundh, "Fourier series for analysis of temporal sequences of satellite sensor imagery," *Int. J. Remote Sens.*, vol. 15, no. 18, pp. 3735–3741, Dec. 1994.

[24] U. Hoefer, J. Frank, and M. Fleischer, "High temperature Ga_2O_3-gas sensors and SnO_2-gas sensors: A comparison," *Sens. Actuators B, Chem.*, vol. 78, nos. 1–3, pp. 6–11, Aug. 2001.

[25] F. Ridzuan and W. M. N. Wan Zainon, "A review on data cleansing methods for big data," *Proc. Comput. Sci.*, vol. 161, pp. 731–738, Jan. 2019.

[26] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*. Burlington, MA, USA: Morgan kaufmann, 2022.

[27] H. Kile and K. Uhlen, "Data reduction via clustering and averaging for contingency and reliability analysis," *Int. J. Electr. Power Energy Syst.*, vol. 43, no. 1, pp. 1435–1442, Dec. 2012.

[28] L. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," in *Proc. 22nd Int. Conf. Distrib. Comput. Syst. Workshops*, 2002, pp. 575–578.

[29] Y. Dodge, D. Cox, and D. Commenges, *The Oxford Dictionary of Statistical Terms*. Oxford, U.K.: Oxford Univ. Press on Demand, 2003.

[30] S. Smyl and K. Kuber, "Data preprocessing and augmentation for multiple short time series forecasting with recurrent neural networks," in *Proc. 36th Int. Symp. Forecasting*, 2016, pp. 1–13.

[31] F. Virili and B. Freisleben, "Nonstationarity and data preprocessing for neural network predictions of an economic time series," in *Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Networks. IJCNN. Neural Comput., New Challenges Perspect. New Millennium*, vol. 5, Jul. 2000, pp. 129–134.

[32] N. Passalis, J. Kanniainen, M. Gabbouj, A. Iosifidis, and A. Tefas, "Forecasting financial time series using robust deep adaptive input normalization," *J. Signal Process. Syst.*, vol. 93, no. 10, pp. 1235–1251, Oct. 2021.

[33] S. Khan, M. A. Alam, N. R. S. Ram, K. Mirza, and V. M. Chowdary, "Noise reduction of time-series satellite data using various de-noising algorithms," *Int. J. Tech. Res. Sci.*, no. 3, pp. 55–69, Aug. 2020.

[34] J. A. Cortés-Ibáñez, S. González, J. J. Valle-Alonso, J. Luengo, S. García, and F. Herrera, "Preprocessing methodology for time series: An industrial world application case study," *Inf. Sci.*, vol. 514, pp. 385–401, Apr. 2020.

[35] D. Puzzovio, "Surface interaction mechanisms in metal-oxide semiconductors for alkane detection," thesis, Universita Degli Studi Di Ferrara, Ferrara, Italy, 2008.

[36] L. M. Miloshevich, A. Paukkunen, H. Vömel, and S. J. Oltmans, "Development and validation of a time-lag correction for Vaisala radiosonde humidity measurements," *J. Atmos. Ocean. Technol.*, vol. 21, no. 9, pp. 1305–1327, Sep. 2004.

[37] Y. Tian, J. Yan, Y. Zhang, T. Yu, P. Wang, D. Shi, and S. Duan, "A drift-compensating novel deep belief classification network to improve gas recognition of electronic noses," *IEEE Access*, vol. 8, pp. 121385–121397, 2020.

[38] E. M. Taguem, L. Mennicken, and A.-C. Romain, "Quantile regression with a metal oxide sensors array for methane prediction over a municipal solid waste treatment plant," *Sens. Actuators B, Chem.*, vol. 334, May 2021, Art. no. 129590.

[39] C. Xie, L. Chao, Y. Qin, J. Cao, and Y. Li, "Using a stochastic forest prediction model to predict the hazardous gas concentration in a one-way roadway," *AIP Adv.*, vol. 10, no. 11, Nov. 2020, Art. no. 115206.

[40] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[41] S. Di Carlo and M. Falasconi, *Drift correction methods for gas Chem. sensors Artif. olfaction systems: Techn. challenges*. INTECH Open Access Publisher, 2012.

[42] Q. Liu, X. Hu, M. Ye, X. Cheng, and F. Li, "Gas recognition under sensor drift by using deep learning," *Int. J. Intell. Syst.*, vol. 30, no. 8, pp. 907–922, 2015.

[43] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, "A comparative study of classification and regression algorithms for modelling students' academic performance," in *Proc. Int. Educ. Data Mining Soc.*, 2015, pp. 1–4.

[44] P. Moeskops, J. M. Wolterink, B. H. Van Der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum, "Deep learning for multi-task medical image segmentation in multiple modalities," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 478–486.

[45] J. Ji, X. Chen, C. Luo, and P. Li, "A deep multi-task learning approach for ECG data analysis," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Mar. 2018, pp. 124–127.

[46] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Deep multi-task multi-channel learning for joint classification and regression of brain status," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 3–11.

[47] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, pp. 1–8.

[48] C. Shin, S. Joo, J. Yim, H. Lee, T. Moon, and W. Rhee, "Subtask gated networks for non-intrusive load monitoring," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 1150–1157.

[49] A. Huss, "Hybrid model approach to appliance load disaggregation," Ph.D. dissertation, KTH Royal Inst. Technol. Stockholm, Stockholm, Sweden, 2015.

[50] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ.*, Nov. 2015, pp. 55–64.

[51] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[52] X. Zhao, Z. Wen, X. Pan, W. Ye, and A. Bermak, "Mixture gases classification based on multi-label one-dimensional deep convolutional neural network," *IEEE Access*, vol. 7, pp. 12630–12637, 2019.

[53] S. De Vito, E. Esposito, M. Salvato, O. Popoola, F. Formisano, R. Jones, and G. Di Francia, "Calibrating chemical multisensory devices for real world applications: An in-depth comparison of quantitative machine learning approaches," *Sens. Actuators B, Chem.*, vol. 255, pp. 1191–1210, Oct. 2018.

[54] M. Capó, A. Pérez, and J. A. Lozano, "An efficient approximation to the *K*-means clustering for massive data," *Knowl. Based Syst.*, vol. 117, pp. 56–69, Feb. 2017.

[55] G. Papamakarios, T. Pavlakou, and I. Murray, "Masked autoregressive flow for density estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[57] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[58] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Res.*, vol. 30, no. 1, pp. 79–82, Dec. 2005.

JEONGYUN HAN received the B.S. and M.S. degrees in mathematics and computer education from Busan National University of Education, Busan, South Korea, in 2006 and 2013, respectively, and the Ph.D. degree in data science from Seoul National University, Seoul, South Korea, in 2019. From 2006 to 2017, he was an Elementary School Teacher with Ulsan Metropolitan Office of Education Korea. From 2020 to 2021, he was a Research Professor with the Korea Advanced Institute of Science and Technology (KAIST), South Korea. He is currently an Associate Research Fellow with the Future Education Research Division, Korean Educational Development Institute (KEDI). His general research interests include building a framework for the effective use of data and constructing a data pipeline for practical application.

MOONJUNG EO received the B.S. and M.S. degrees from Seoul National University, Seoul, South Korea, in 2011 and 2013, respectively, where she is currently pursuing the Ph.D. degree with the Department of Intelligence and Information. Her current research interests include deep learning, such as implicit and explicit regularization methods for generalization and network compression methods.

WONJONG RHEE (Fellow, IEEE) received the B.S. degree in electrical engineering from Seoul National University, Seoul, South Korea, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1998 and 2002, respectively. From 2001 to 2003, he was with ArrayComm as a Research Staff. From 2004 to 2013, he was with ASSIA Inc., as a Founding Member and later as an ASSIA Fellow. He is currently an Associate Professor with the Department of Intelligence and Information, Seoul National University. His general research interests include data science and machine learning, where information theory, optimization, and signal processing are utilized.

• • •