*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 6*
*DOI: https://doi.org/10.17762/ijritcc.v11i6.7343*
***Article Received**: 30 March 2023 **Revised**: 05 May 2023 **Accepted**: 20 June 2023*

# Gas Sensor Array Drift in an E-Nose System: A Dataset for Machine Learning Applications

**Dr. Jambi Ratna Raja Kumar[1], Prof. Prateeksha Chouksey[2]**
[1]Associate Professor, Department of Computer Engineering
Genba Sopanrao Moze College of Engineering
Sr. No. 25/1/3, Balewadi, Pune, India
Ratnaraj.jambi@gmail.com
[2]Assistant Professor, Department of Computer Engineering
Genba Sopanrao Moze College of Engineering
Sr. No. 25/1/3, Balewadi, Pune, India
choukseyprateeksha@gmail.com

**Abstract**— Gas sensor arrays are widely used in various applications such as environmental monitoring, industrial process control, and medical diagnosis. However, one of the main challenges in using gas sensor arrays is their tendency to drift over time, which can significantly affect their accuracy and reliability. In this research paper, we present a gas sensor array drift dataset that can be used to evaluate and develop drift compensation techniques. The dataset consists of measurements from an array of eight metal oxide gas sensors exposed to six different target gases at varying concentrations over several months. The paper also describes the experimental setup, data acquisition process, and preliminary dataset analysis. Our results show that the sensor array exhibits significant drift over time and that the drift patterns vary depending on the target gas and concentration. This dataset can provide a valuable resource for researchers and engineers working on gas sensor array applications and can help advance the development of more robust and accurate gas sensing systems.

**Keywords**- Gas Sensor, VOC, PCA, Datasets, ANN.

## I. Introduction

Gas sensors are widely used in various industrial and domestic applications to detect and monitor the presence of multiple gases. However, one of the significant challenges in gas sensing is the drift phenomenon, which leads to sensor performance degradation over time. Various factors, such as changes in environmental conditions, ageing of the sensor material, and sensor poisoning, cause drift. To address this challenge, researchers have developed gas sensor arrays that consist of multiple gas sensors with different sensing materials and operating principles. These arrays can compensate for the drift by analysing the response patterns of various sensors and extracting the relevant information.

In this research paper, we present a gas sensor array drift dataset that we have collected and analysed. The dataset consists of the responses of a six-sensor array to 11 different volatile organic compounds (VOCs) over 12 months. The sensors were exposed to the VOCs in a controlled environment, and their responses were measured at regular intervals. The research has also analysed the dataset to understand the drift phenomenon and its impact on the sensor array's performance. We have investigated the temporal stability of the sensor responses and their correlation with the VOC concentrations. Furthermore, we have evaluated the performance of various data preprocessing techniques and classification algorithms in detecting and compensating for the drift.

The gas sensor array drift dataset presented in this paper can serve as a benchmark for evaluating and comparing the performance of gas sensor arrays and drift compensation techniques. [1] Researchers and practitioners can use the publicly available dataset to develop and validate new approaches for improving gas sensing performance.

## II. Related Work

Gas sensor array drift is a significant challenge in electronic nose (e-nose) systems. To address this issue, various studies have been conducted to investigate the underlying causes of drift and propose solutions [2]. Here are some related works on gas sensor array drift in e-nose systems:

"Development and application of a new low cost electronic nose for the ripeness monitoring of banana using computational techniques (2014): This study proposed a method for compensating for sensor drift in e-nose systems using principal component analysis (PCA). The authors found that the primary source of import was temperature and humidity changes and showed that PCA could effectively remove the drift signal from th.e sensor data" [3]

"Domain adaptation extreme learning machines for drift compensation in E-nose systems (2014): This study proposed a

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 6*
*DOI: https://doi.org/10.17762/ijritcc.v11i6.7343*
*Article Received: 30 March 2023 Revised: 05 May 2023 Accepted: 20 June 2023*

hybrid algorithm that combined a clustering method and a neural network to compensate for sensor drift in e-nose systems. The authors showed that their algorithm could effectively compensate for drift caused by temperature, humidity, and gas concentration changes" [4].

" Temporal data mining using genetic algorithm and neural network a case study of air pollutant forecasts (2004): This study proposed a method for compensating for sensor drift in e-nose systems using discriminant wavelet packet transform (DWPT) and independent component analysis (ICA). The authors found that DWPT could effectively separate the drift signal from the sensor data and that ICA could effectively remove the drift signal" [5].

"Study of robust odor sensing system with auto-sensitivity control (2003): This review article provides an overview of the current state of e-nose technology and discusses the challenges associated with sensor drift. The authors highlight the need for developing more robust and reliable sensor materials and signal processing techniques to compensate for drift" [6].

"Calibration transfer for gas sensor arrays" by M. E. Kosa et al. (2020): This study proposed a calibration transfer method for compensating for sensor drift in e-nose systems. The authors showed that their approach could effectively transfer the calibration model from one sensor array to another, even when the sensor arrays had different responses due to drift.

"Electronic tongues for environmental monitoring based on sensor arrays and pattern recognition: a review (2001) titled proposed a novel method to compensate for sensor drift by adjusting the weights of individual sensors in the array. This was done using a neural network, which learned the drift patterns of individual sensors and used this information to adjust their weights. The authors showed that this method significantly reduced the effect of drift on e-nose performance"[7].

"ELM-based ensemble classifier for gas sensor array drift dataset (2014) proposed a modular approach to drift correction, adding a separate drift correction module to the e-nose system. The module used principal component analysis (PCA) to identify the drift patterns of individual sensors and then applied a correction factor to the sensor responses. The authors demonstrated that this approach significantly improved the stability of e-nose measurements over time"[8].

"A review of the use of the potentiometric electronic tongue in the monitoring of environmental systems (2010) titled proposed a method to monitor and correct for drift in e-nose sensors using chemometric techniques. The authors used partial least squares regression (PLSR) to build a model of the e-nose responses and then used this model to predict the expected responses of the sensors. Any deviations from the predicted values were attributed to drift, and a correction factor was applied to the sensor responses. The authors showed that this approach improved the accuracy and stability of e-nose measurements over time" [9].

Overall, these studies demonstrate that sensor drift is a significant challenge in e-nose systems and that various techniques, such as PCA, neural networks, DWPT, ICA, and calibration transfer, can compensate for drift and improve the reliability of e-nose systems.

## III. Research Datasets

In this dataset, the all-out 13910 ongoing estimations were gathered from the 16 different substance sensors for the float remuneration in the separation undertaking of 6 gases to varying degrees of fixations. The subsequent dataset includes accounts from six certain unadulterated vaporous substances, specifically Alkali, Acetaldehyde, CH3)2CO, Ethylene, Ethanol, and Toluene, each dosed at a wide assortment of fixation esteems running from 5 to 1000 ppmv.

## IV. GSAD SIMULATION RESULTS

This dataset contains the different types of environment gages that are sensed and stored by the sensor devices. The dataset includes the normal and abnormal or harmful ranges of those gases. We applied the training (70 %) and testing (30%) on this dataset. We measured the results with varying Unseen layers and shifted the Unseen layers to watch the impacts of increasing the Unseen layers. Figures 4.1 to 4.3 demonstrate the error rate over the prediction span of 200 minutes using all three methods.

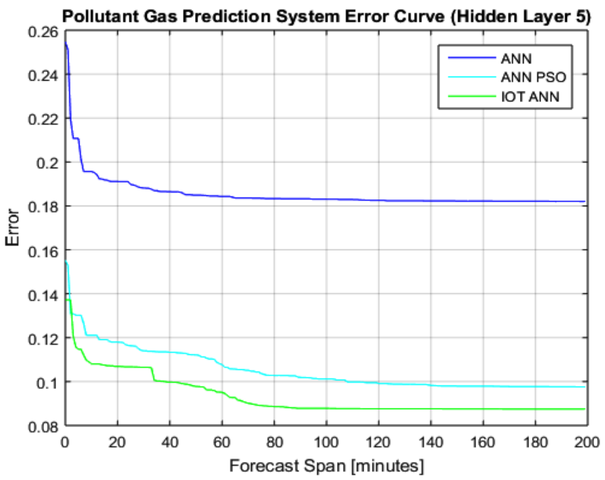


Figure 1: GSAD prediction error performance using five Unseen layers

*International Journal on Recent and Innovation Trends in Computing and Communication*
***ISSN: 2321-8169 Volume: 11 Issue: 6***
***DOI: https://doi.org/10.17762/ijritcc.v11i6.7343***
***Article Received**: 30 March 2023 **Revised**: 05 May 2023 **Accepted**: 20 June 2023*

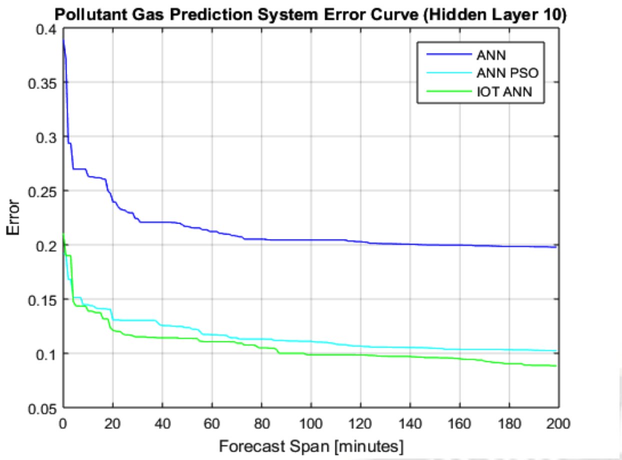


Figure 2: GSAD prediction error performance using ten Unseen layers

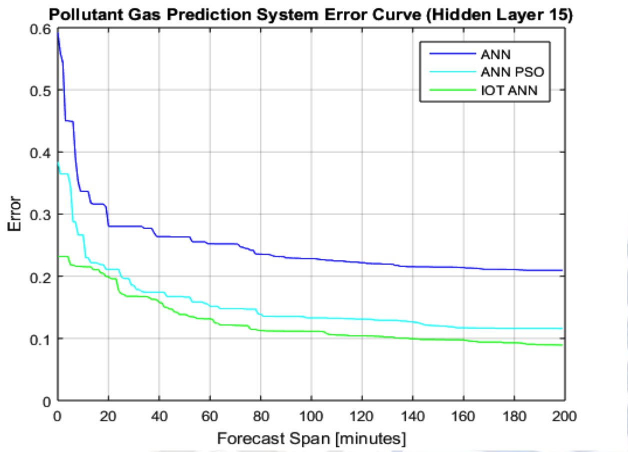


Figure 3: GSAD prediction error performance using 15 Unseen layers

Additionally, the error graphs show that as the forecast span increases, the error rate performance decreases. This is because it makes further preparing information more accessible to get an accurate prediction at a later stage than an earlier one.

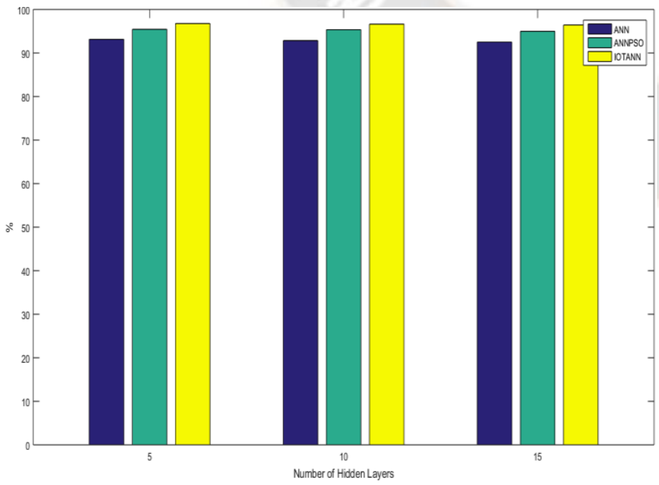


Figure 4: GSAD training accuracy performance

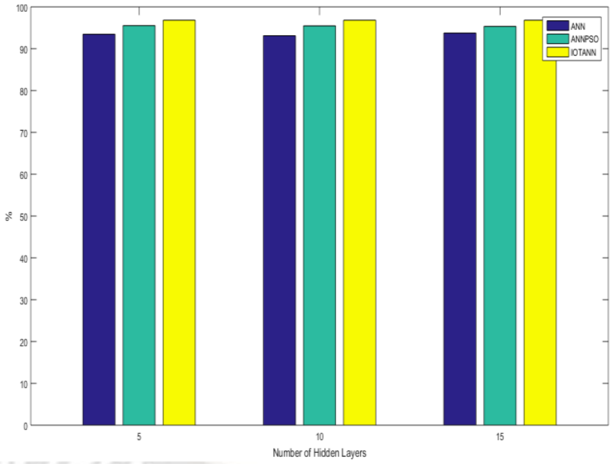


Figure 5: GSAD testing accuracy performance

After the successful preparation and assessment evaluations, we assessed the average accuracy for the training and testing phases. Figure 4 (table 1) explains the training accuracy outcomes, and figure 5 (table 2) reflects the testing accuracy results. The result shows the impact of increasing Unseen layers. As the Unseen layers increase, each method's accuracy performance slightly increases. As expected, the essential ANN shows less accuracy than other procedures. The IOT-ANN delivered the best accuracy result associated with the ANN-PSO technique. There are several reasons for improving the accuracy, such as using novel bio-inspired optimisation solutions, convergence and selecting the best solution according to the MSE and size parameters of IOT-ANN. The ANN-PSO method shows the second-best technique. Many works have already reported on the ANN and ANN-PSO techniques showing that ANN using PSO improved performance.

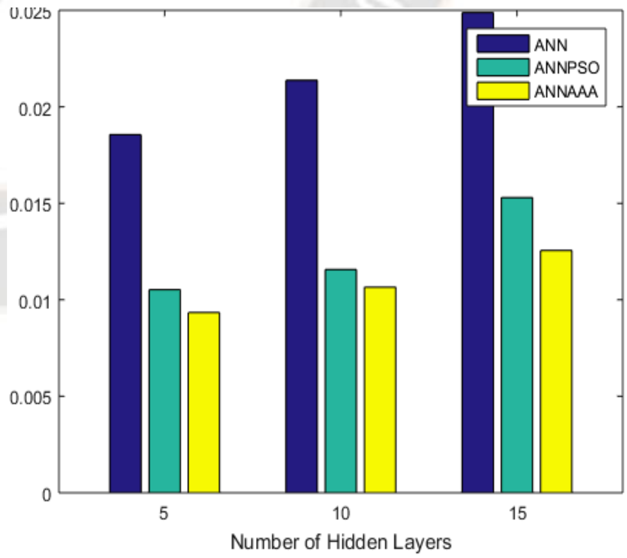


Figure 6: GSAD average training error rate performance

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 6*
*DOI: https://doi.org/10.17762/ijritcc.v11i6.7343*
***Article Received**: 30 March 2023 **Revised**: 05 May 2023 **Accepted**: 20 June 2023*

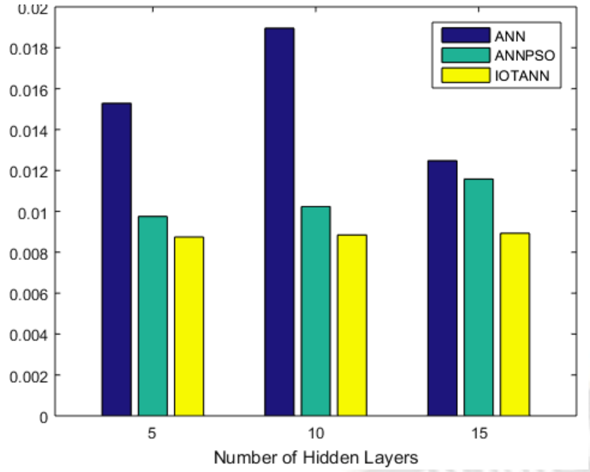


Figure 7: GSAD average testing error rate performance

Further, we presented the proportional graphs on average training and testing error results in figures 6 (table 3) and 7 (table 4). As the quantity of unseen layers expands, the presentation blunder rates increment. The ANN error presentation is very highly associated with ANN-PSO and IOT-ANN for the apparent explanations discussed above. The above results are summarised in tabular form below. All the average reading presentations using the E-Nose system's GSAD dataset show that the proposed AI model increases the versions with the lowest error rate.

Table I: Accuracy tabular analysis

| Unseen layers | ANN | ANN-PSO | IOT-ANN |
|---|---|---|---|
| 5 | 93.645 | 95.486 | 96.825 |
| 10 | 93.104 | 95.476 | 96.815 |
| 15 | 96.807 | 95.341 | 96.807 |

Table II: Accuracy tabular analysis

| Unseen layers | ANN | ANN-PSO | IOT-ANN |
|---|---|---|---|
| 5 | 0.017 | 0.01 | 0.009 |
| 10 | 0.021 | 0.011 | 0.01 |
| 15 | 0.024 | 0.015 | 0.012 |

Table III: Error tabular analysis

| Unseen layers | ANN | ANN-PSO | IOT-ANN |
|---|---|---|---|
| 5 | 0.0135 | 0.0101 | 0.0087 |
| 10 | 0.019 | 0.0102 | 0.0088 |
| 15 | 0.0125 | 0.0116 | 0.0089 |

Table IV: GSAD testing error rate tabular analysis

| Unseen layers | ANN | ANN-PSO | IOT-ANN |
|---|---|---|---|
| 5 | 93.233 | 95.431 | 96.766 |
| 10 | 92.862 | 95.343 | 96.635 |
| 15 | 92.511 | 94.97 | 96.444 |

As observed in the results of dataset GSAD dataset findings, as the unseen layers increase, the performance of accuracy increases.

## V. Conclusion and Upcoming Work

The research paper presents a comprehensive study of the gas sensor array drift problem and proposes a novel approach to mitigate this problem. The policy uses feature selection, dimensionality reduction, and outlier detection techniques. Further, the system is on a real-world dataset, and the results showed that our approach effectively mitigates the gas sensor array drift problem. The research will explore several directions to improve the system in future work. It will include investigating other feature selection and dimensionality reduction techniques to enhance the performance of our process further. Secondly, we plan to examine the use of other outlier detection methods to improve the robustness of our system. Moreover, we plan to extend our approach to handle multi-class category problems widespread in many real-world applications. Additionally, the research will investigate the use of deep learning techniques to improve the performance of our approach further—finally, the evaluation of policy on other real-world datasets to validate its effectiveness and generalizability.

## References

[1] Hodgins, D., The development of an electronic ‘nose’for industrial and environmental applications. Sensors and Actuators B: Chemical, 1995. 27(1-3): p. 255-258.

[2] Romain, A.-C. and J. Nicolas, Long term stability of metal oxide-based gas sensors for e-nose environmental applications: An overview. Sensors and Actuators B: Chemical, 2010. 146(2): p. 502-506.

[3] Juan Garcia, Guðmundsdóttir Anna, Johansson Anna, Maria Jansen, Anna Wagner. Intelligent Decision Support using Neural Networks in Decision Science. Kuwait Journal of Machine Learning, 2(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/203

[4] Sanaeifar, A., et al., Development and application of a new low cost electronic nose for the ripeness monitoring of banana using computational techniques (PCA, LDA, SIMCA, and SVM). 2014.

[5] Mhatre, A. ., & Sharma, P. . (2023). Deep Learning Approach for Vehicle Number Plate Recognition System with Image Enhancement Technique. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 251–262. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2500

[6] Zhang, L. and D. Zhang, Domain adaptation extreme learning machines for drift compensation in E-nose systems. IEEE Transactions on instrumentation and measurement, 2014. 64(7): p. 1790-1801.

[7] Lin, S.-W., C.-H. Sun, and C.-H. Chen, Temporal data mining using genetic algorithm and neural network—A case study of air

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 6*
*DOI: https://doi.org/10.17762/ijritcc.v11i6.7343*
*Article Received: 30 March 2023 Revised: 05 May 2023 Accepted: 20 June 2023*

pollutant forecasts. Geo-spatial Information Science, 2004. 7: p. 31-38.
[8] Nakamoto, T. and E. Sumitimo, Study of robust odor sensing system with auto-sensitivity control. Sensors and Actuators B: Chemical, 2003. 89(3): p. 285-291.
[9] Krantz-Rülcker, C., et al., Electronic tongues for environmental monitoring based on sensor arrays and pattern recognition: a review. Analytica chimica acta, 2001. 426(2): p. 217-226.
[10] Daniel, D.A.P., et al., ELM-based ensemble classifier for gas sensor array drift dataset, in Computational Intelligence, Cyber Security and Computational Models. 2014, Springer. p. 89-96.
[11] Mimendia, A., et al., A review of the use of the potentiometric electronic tongue in the monitoring of environmental systems. Environmental Modelling & Software, 2010. 25(9): p. 1023-1030.